Donald E. Farina
Section 5-96 Radar Section
Moorestown N.J. Bldg 101-2

5

Digitized by the Internet Archive
in 2017 with funding from
Kahle/Austin Foundation

https://archive.org/details/principlesofrada00mass

PRINCIPLES OF RADAR

PRINCIPLES OF RADAR

By Members of the Staff of the Radar School
Massachusetts Institute of Technology

Third Edition by

J. Francis Reintjes
ASSISTANT PROFESSOR OF COMMUNICATIONS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Godfrey T. Coate
FORMERLY ASSISTANT PROFESSOR OF
ELECTRICAL ENGINEERING
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

A publication of the Technology Press
Massachusetts Institute of Technology

McGRAW-HILL BOOK COMPANY, INC.

New York Toronto London

1952

PRINCIPLES OF RADAR

Copyright, 1944, 1946, 1952, by Massachusetts Institute of Technology, Cambridge, Massachusetts. Printed in the United States of America. All rights reserved. Subject to a nonexclusive royalty-free license to the Government of the United States of America. This book, or parts thereof, may not be reproduced in any form without permission of the publishers.

Library of Congress Catalog Card Number: 52-5996

THE MAPLE PRESS COMPANY, YORK, PA.

FOREWORD

Just prior to the entry of the United States into the Second World War, the Massachusetts Institute of Technology, in cooperation with the Armed Services, established a Radar School at M.I.T. for the purpose of training selected groups of Army and Navy officers in the principles and techniques of radar. Throughout the war years, the School provided full-time programs of instruction in the basic principles of radar and in the operating features of specific radar sets. After the war, the Radar School continued to operate for approximately three years as a United States Naval School, Electronics Engineering, for selected Naval officers studying to become electronics specialists. Since discontinuance of the School in 1949, a limited program of instruction in radar has been offered at M.I.T. in the form of a fourth-year elective subject in electrical engineering.

Soon after the establishment of the Radar School in 1941, preparation of the first edition of *Principles of Radar* was undertaken to provide a reference text for use in the program of study. The various chapters were written, for the most part, by members of the Radar School lecture staff and followed approximately the content and presentation of the lectures on basic radar principles. The book was first issued to Radar School students as a classified document in September, 1944.

Because of the rapid advances being made in the field of radar at the time, it was decided to begin a revision of the book almost immediately. The primary objectives of this revision were to improve the clarity of the presentations and to bring the book up to date. A complete reorganization of the subject matter was also considered advisable in order to obtain a more unified treatment, but this plan was fulfilled only in part because of the urgent need for the revised edition. The order of chapters established in the first edition was preserved, but several chapters were rewritten principally to bring about better integration between parts of the book. The revised edition was prepared from material provided by a group of Radar School staff members. The manuscript was completed almost coincidently with the termination of the war, and clearance to release the material for general publication was obtained toward the end of 1946.

The extensive revision originally considered for the second edition has been carried out in the present volume. This new edition, prepared by Prof. J. F. Reintjes and Prof. G. T. Coate of the Department of Electrical Engineering, M.I.T., benefits by the efforts of the many persons

who contributed in one way or another to the earlier volumes. Preparation of the manuscripts for the first and second editions was carried out under the supervision of the Curriculum Committee of the Radar School, of which W. H. Radford was chairman. Other members of the Committee who were active in this matter were: J. M. Henry, W. J. Jackson, H. C. Kendall, and H. J. Zimmermann.

Principal contributors to the first edition were: W. L. Barrow (Radar School Director, 1941–1943), J. C. Batchelor, A. Coale, G. T. Coate, A. F. Coleman, E. W. Cowan, Jr., M. Eager, S. K. Haynes, A. H. Howell, Jr., E. W. Kimbark (Editor), J. P. LaSalle, R. D. Spence, and R. L. Wooley.

Principal contributors to the second edition were: G. E. Bowler, G. T. Coate (Assistant Editor), W. P. Cunningham, M. Eager, M. W. Essigmann, L. W. Evans, R. A. Galbraith, S. K. Haynes, E. W. Keller, R. L. Mattingly, M. S. McIlroy, J. F. Reintjes (Editor), A. Sheingold, R. D. Spence, J. N. Steele, M. A. Townsend, and H. J. Zimmermann.

In addition to the above contributors many other Radar School staff members rendered valuable assistance by reviewing, criticizing, and rewriting portions of the manuscripts. Acknowledgment of the work of all who participated in the project is hereby made, and thanks are expressed for their efforts.

Carlton E. Tucker, Executive Officer
Department of Electrical Engineering
Massachusetts Institute of Technology

Cambridge, Mass.
May, 1952

PREFACE

This volume, like the two preceding editions, deals with the basic principles of pulse radar. It differs from the other editions, however, in a number of ways. The order of presentation of the subject matter has been completely reorganized, and nearly all the material has been rewritten to introduce improved techniques of presentation and to bring about better continuity of articles and chapters. The subject matter has also been expanded by approximately twenty per cent, and more than half of the illustrations are new.

The book is intended primarily for those interested in the basic concepts and techniques of radar. A good deal of the material has been made sufficiently general, however, to render it useful to those working in other fields employing similar techniques. For most efficient use of the book, the reader should have prior knowledge of vacuum-tube behavior and the basic principles of direct- and alternating-current circuits. Ability to carry out simple differentiation and integration is also assumed.

In the frequency domain, radar components divide themselves naturally into low-frequency and high-frequency groups, and it is on this basis that the subject matter is organized. Thus the fore part of the book, which includes the first six chapters, pertains, in general, to pulse circuits and their application to radar modulators, indicators, and receivers. Synchro and servomechanism control circuits are also worked into their logical place in this section. The latter part of the book is devoted to the radio-frequency aspects of radar. Basic concepts are developed in the chapters on transmission lines, waveguides, and resonant cavities, and the techniques of their employment in radar are presented in the remaining chapters.

It is recognized that the concepts presented in a number of articles evolved, either directly or indirectly, from the ideas and thoughts of contributors to other editions of the book. Acknowledgment of the influence of these contributors on the present manuscript is gratefully made. Some of the material from the preceding edition has also been retained essentially intact. In this category is the chapter on propagation, which was originally written for the second edition by Prof. Henry J. Zimmermann of the Department of Electrical Engineering, M.I.T.

In carrying out the project, the authors received the support and encouragement of Dean Harold L. Hazen, and of Prof. Carlton E. Tucker, Executive Officer of the Department of Electrical Engineering and

Director of the Radar School from 1943 to 1949. Professor William H. Radford also took an active part in guiding the project to completion, and Prof. Marcy Eager rendered valuable assistance through his criticisms and proofreading of the manuscript. The cooperation received from Frederick G. Fassett, Jr. and Prof. Gordon S. Brown of the Technology Press is gratefully acknowledged.

Most of the new illustrations were prepared by the M.I.T. Illustration Service. Appreciation is expressed to everyone in this group who participated in the project, and in particular to Gloria B. Williams who was primarily reponsible for the work. Appreciation is also expressed to Dorothy Young who typed nearly all the manuscript and assisted in the editing. Thanks are extended to the General Electric Company, Radio Corporation of America, and the Bell Telephone Laboratories for their cooperation in supplying photographs.

J. Francis Reintjes
Godfrey T. Coate

CAMBRIDGE, MASS.
May, 1952

CONTENTS

FOREWORD v

PREFACE vii

CHAPTER I. *Pulsed Radar* 1

1. Radar Determination of Range 2
2. Radar Direction Indication 7
3. Data Presentation 12
4. Discernment of Signals and Noise 16
5. Radar Range Equation 18
6. Limitations of Radar 23
7. Radar Frequencies 32
8. Types of Radar Equipment 39
9. Block Diagram of a Radar Set 46

CHAPTER II. *Timing Circuits* 54

A. CLIPPING CIRCUITS 54
1. Clipping with Diodes 54
2. Clipping with Amplifiers 56
3. Clipping with Cathode Followers 61

B. R-C CIRCUITS 66
4. Transients in *R-C* Circuits 67
5. Saw-tooth Generator 69
6. Decoupling Filters and Bias Circuits 73
7. *R-C* Peaker 75
8. Sloping Wavefront Applied to Peaker 78
9. *R-C* Coupling Circuit 79
10. Signal Bias and Clampers 82
11. Effect of Grid Return 83
12. Step Generator 86
13. Free-running Plate-coupled Multivibrator 92
14. Driven Plate-coupled Multivibrator 97
15. Cathode-coupled Multivibrator 99
16. Phantastron Delay Circuit 104
17. Tube and Wiring Capacitances 110

C. R-L-C CIRCUITS 114
18. Transients in Parallel *R-L-C* Circuits 115
19. Ringing Circuit 120

20. Pulse Transformers 123
21. Blocking Oscillator 134
22. Traveling Waves on Artificial Transmission Lines 139
23. Switching Operations in Transmission-line Circuits 143
24. Reflections at Line Terminals 145
25. Applications of Artificial Transmission Lines 150

CHAPTER III. *Modulators* 153

A. PULSE FORMATION BY NETWORKS AND SWITCHES 154
1. Charging an Open-end Line 154
2. Pulse Formation by Discharge of an Open-end Line 158
3. Pulse Formation by a Shorted-end Line 159
4. Pulse-forming Networks 160
5. Characteristics of Loads for Pulse-forming Networks 164
6. Switching Devices for Discharging Pulse-forming Networks 166

B. LINE-PULSING MODULATORS 169
7. Charging Methods 170
8. Rotary-spark-gap Modulator 176
9. Modulator with Saturable-core Inductor and Inductive-impulse Charging 183

C. DRIVER-POWER-AMPLIFIER MODULATORS 186
10. *R-C*-coupled Power Amplifier 187
11. Transformer-coupled Power Amplifier 196
12. Bootstrap Drive 200
13. Line-controlled Blocking-oscillator Driver 204
14. Delay-line-controlled Driver 209
15. Driver Using Shorted-end Line 210
16. Blocking-oscillator Modulator 214

CHAPTER IV. *Indicators* 215

A. CATHODE-RAY TUBES 215
1. Electrostatic Cathode-ray Tubes 216
2. Magnetic Cathode-ray Tubes 221
3. Designations and Ratings of Cathode-ray Tubes 225
4. Data Presentation by Means of Cathode-ray Tubes 225

B. SWEEP-VOLTAGE CIRCUITS 227
5. Basic Type A Indicator 227
6. Methods of Improving the Linearity of Sweep Voltages 229
7. Phase-inverter Circuits 234
8. Delayed Sweeps and Expanded Sweeps 239
9. Indicators Used with Antenna Switching 240
10. Type J Presentation 242

11. Ringing Circuit for Type J Presentation 244
12. Crystal-controlled Oscillator for Type J Presentation 248

C. SWEEP-CURRENT CIRCUITS 252
13. Basic Magnetic-deflection Indicator 252
14. Current-Voltage Relations for Deflection Coil 253
15. Grid Voltage for Deflection-coil Driver Tube 255
16. *R-C* Trapezoidal-voltage Generator 258
17. Rectangular Grid-voltage Pulse for Driver Tube 260
18. Feedback-amplifier Sweep Circuit 262
19. Azimuth Sweep Circuit for Type B Presentation 265
20. Rotating-coil Plan-position Indicator 267
21. Methods of Positioning the Sweep Trace 269
22. Fixed-coil Plan-position Indicator 271

CHAPTER V. *Synchros and Servomechanisms* 277

A. SYNCHROS 277
1. Construction and Connection of Synchro Motors and Generators 277
2. Operation of Generator—Electrical Zero 280
3. Operation of Motor 284
4. Torque-Angle Curve 287
5. Oscillation Dampers 288
6. Synchro Control Transformer 289
7. Differential Synchro 292
8. Designations of Synchros 295
9. Synchro Capacitors 296
10. The Electrical-zero Autotransformer 298
11. Geared Synchro Systems 298
12. Applications of Synchros 300

B. SERVOMECHANISMS 305
13. Components of Servo Systems 306
14. Servo Using Ward-Leonard Control System 308
15. Direct-current Servo Motor 309
16. Phase-sensitive Rectifiers 312
17. Amplidyne Servomechanism 316
18. Two-phase Induction Motor 319
19. Servomechanisms Using Two-phase Induction Motor 322
20. Transient Oscillations in Servomechanisms 324
21. Sustained Oscillations in Servomechanisms 325
22. Antihunt Devices 328
23. Applications of Servomechanisms 336

CHAPTER VI. *Receivers* 338

A. GENERAL PRINCIPLES 338
1. Block Diagram 338
2. Sine-wave Analysis 341

3. Sine-wave Components of Video Pulses 343
4. Sine-wave Components of Carrier-frequency Pulses 346
5. Sine-wave Response of *R-C* Circuits 349
6. Sine-wave Response of Parallel *R-L-C* Circuit 352
7. Reproduction of Video Pulses 356
8. Reproduction of Carrier-frequency Pulses 359
9. Linear Equivalent Circuits for Tubes 364
10. Grid Loading Caused by Cathode-lead Inductance 367
11. Transit-time Grid Loading 369
12. Silicon Crystals 371
13. Characteristics of Noise 376
14. Noise in Linear Passive Circuits 379
15. Antenna Noise 383
16. Vacuum-tube and Crystal Noise 384
17. Noise Figure 392
18. Receiver Frequencies and Bandwidths 396

B. COMPONENT CIRCUITS 401

19. Typical Video Amplifier 401
20. Video-pulse Amplification 404
21. Compensated Video Amplifiers 409
22. Single-tuned Intermediate-frequency Amplifier 412
23. Double-tuned Coupling Circuits 419
24. Stagger-tuned Amplifier 424
25. Intermediate-frequency Input Stages 427
26. Diode Detectors 432
27. Triode and Pentode Detectors 439
28. Crystal Mixers 441
29. Vacuum-tube Mixers 449
30. Automatic-frequency-control Circuits 452
31. Anti-interference Circuits and Signal Discernibility 461

CHAPTER VII. *Radio-frequency Transmission Lines* 468

1. Types of Transmission Lines 469
2. Traveling Waves 471
3. Lines Terminated in a Short Circuit 477
4. Lines Terminated in an Open Circuit and in a Reactance 483
5. Lines Terminated in a Resistance 486
6. Lines Terminated in a Complex Impedance 490
7. Graphical Determination of Reflection Coefficient 493
8. Input Impedance of Transmission Lines 495
9. Transmission-line Charts 499
10. Input Admittance 505
11. Transient and Steady-state Operation of Transmission Lines 506
12. Power and Power Capacity 510

13. Losses and Attenuation 514
14. Impedance of Lines Having Small Losses 521
15. Impedance-changing Devices 523
16. Transmission-line Capacitance 531
17. Transmission-line Inductance 535
18. Transmission-line Resistance 539
19. Characteristic Resistance, Propagation Speed, and Attenuation 541

CHAPTER VIII. *Waveguides* 545

1. Applications and Advantages of Waveguides 545
2. Basic Electromagnetic-wave Concepts—The Parallel-strip Transmission Line 547
3. Types of Electromagnetic Waves 554
4. Normal Reflection of Plane Transverse Waves 556
5. Oblique Reflection of Plane Transverse Waves 560
6. Field Patterns Obtained by Oblique Reflection 564
7. The $TE_{1,0}$ Wave in a Rectangular Waveguide 570
8. Phase Velocity and Group Velocity 577
9. Power and Attenuation 580
10. Higher Order Wave Types in Rectangular Waveguides 587
11. Rectangular-waveguide Dimensions 590
12. Waveguide Fields for Frequencies below Cutoff 591
13. Waveguide Impedances 596
14. Impedance-changing Devices 601
15. Circular Waveguides 608

CHAPTER IX. *Resonant Lines and Cavities* 613

1. Resonant Modes of Networks 613
2. The Quality Factor Q 617
3. Transmission-line Resonators 619
4. Rectangular-waveguide Resonators 625
5. Cavity Resonators 633
6. Impedance of One-entry Resonators 639
7. Coupling to One-entry Resonant Networks 645
8. Equivalent Circuits for Transmission-line Resonators 649
9. Coupling to Transmission-line Resonators 654
10. Transmission Lines Terminated by Resonators 658
11. Two-entry Resonators 665
12. Coupling Devices for High-frequency Resonators 668
13. Parallel Connections at High Frequency 670
14. Coupling Loops 674
15. Wavemeters 677
16. Coupling Probes 682

17. Waveguide-to-coaxial-line Junctions 684
18. Aperture Coupling 689

CHAPTER X. *High-frequency Triode Oscillators* 697

1. Ultrahigh-frequency Triodes 697
2. Sustained Oscillations 703
3. Ultra-audion Oscillators 706
4. Plate-return Oscillators 709
5. Multitube Oscillators 712
6. Analysis of Grid-return Circuits 716
7. Grid-return Oscillators 720
8. Loading of Oscillators 724
9. Output Circuits for Uhf Oscillators 727
10. Effects of Load on Frequency of Oscillation 734
11. Grid-leak Bias in Oscillators 737
12. Pulsed Oscillators 740

CHAPTER XI. *Klystrons and Magnetrons* 745

A. KLYSTRONS 745
1. Two-resonator Klystron 745
2. Reflex Klystrons 750
3. Characteristics of Reflex Klystrons 755
4. Electronic Tuning, Hysteresis, and Load Effects 760

B. MAGNETRONS 763
5. Magnetron Construction and Type Numbers 763
6. Magnetron Resonators 769
7. Electron Motion in Magnetrons 776
8. Magnetron Characteristics 781
9. Performance Charts 792
10. Rieke Diagrams 795
11. Operation and Care of Magnetrons 803

CHAPTER XII. *Radio-frequency Transmitting and Receiving Systems* 806

1. Radio-frequency-system Components 806
2. Choke Joints 810
3. Rotary Joints 812
4. Waveguide Bends, Corners, and Twists 816
5. T Junctions 818
6. Hybrid Junctions 825
7. Directional Couplers 834
8. Attenuators 840
9. Balanced-to-unbalanced Transformers 849
10. Adjustment of Impedance Levels 851
11. Tubes for T-R Devices 857

12. Ultrahigh-frequency T-R Systems 860
13. Transmit-Receive Devices for Coaxial-line Microwave Systems 863
14. Transmit-Receive Devices for Waveguide Systems 866
15. Input Circuits for Ultrahigh-frequency Receivers 871
16. Single-ended Crystal Mixers 876
17. Balanced Crystal Mixers 881
18. Test Equipment 886

CHAPTER XIII. *Antennas* 895

1. The Half-wave Antenna 895
2. Input Impedances of Center-fed Antennas of Arbitrary Length 899
3. Radiation Fields of a Half-wave Antenna 903
4. Analysis of Radiation Fields of a Half-wave Antenna 905
5. Patterns of a Half-wave Antenna; Pattern Interpretation 910
6. Radiation Pattern of an Array of Half-wave Antennas Fed in Phase 913
7. Radiation Pattern of an Array of Half-wave Antennas Fed out of Phase 919
8. Examples of the Use of Factors in Obtaining Antenna Patterns 924
9. Elimination of Back Radiation from Broadside Arrays 929
10. Approximate Method of Determining the Beam Angle of a Broadside Array 933
11. Antenna Connections for Broadside Arrays 934
12. Parasitic Antennas 937
13. Antennas with Parabolic Reflectors 942
14. Feeds for Parabolic Antennas 950
15. The Cosecant-squared Antenna 957

CHAPTER XIV. *Propagation* 959

1. Components of Radiation 960
2. Reflection from a Perfectly Conducting Plane 961
3. Reflection Coefficients 964
4. Effect of Earth Reflections on Range 965
5. Divergence 967
6. Standard Atmospheric Refraction 968
7. Vertical-coverage Diagrams 971
8. Nonstandard Refraction 974
9. Atmospheric Reflections 976
10. Character of Radar Echoes 976

INDEX 979

CHAPTER I

PULSED RADAR

The word *radar*, formed from the italicized letters in *ra*dio *d*etection *a*nd *r*anging, signifies a means of employing radio waves to detect and locate material objects. Location of an object is accomplished by determining the distance and direction from the radar equipment to the object and requires, in general, the measurement of three coordinates—usually range and angles of azimuth and elevation.

Radar detection depends upon the reflection of radio waves from an object. A radar set may be compared with a searchlight and an observer stationed beside the light. A small portion of the light wave from the searchlight strikes an object near by and is reflected. A fraction of the reflected wave returns to the observer and indicates the presence of the object. Similarly, a little of the energy of the radio wave transmitted from the radar set travels to the object, a portion is reflected, and some of the reflected energy returns to the radar set, where it is detected. Advantages of radar over detection systems employing light are:

1. Radar operates over greater ranges—aircraft can be detected more than 100 miles away.

2. Radar operates in any weather, through smoke, fog, rain, or snow, or directly into blinding sun.

3. Radar determines range simply and accurately by measuring the time required for the two-way trip of the radio wave.

Another feature of radar, which is an advantage in wartime applications, is that radio waves are invisible. Thus the enemy must have special equipment in order to discover that a radar set is searching for him.

The reflected waves—called *radio echoes* because of the analogy to sound echoes—enable radar sets to perform a variety of tasks. Applied first in wartime, radar searched the sea for enemy ships and surfaced submarines and kept watch in the sky for aircraft. Radar aided air and sea navigation and directed fighter and bomber aircraft. Radar information was used to aim large guns and antiaircraft weapons, and air-borne radar controlled the dropping of bombs. In peacetime, radar and allied techniques contribute greatly to the safety of air and sea travel. Radar-equipped ships move through fog without danger of colliding with other ships or with ice floes. Despite weather conditions of poor visibility, aircraft can be prevented from colliding in mid-air, and the capacity of air

fields can be greatly increased by radar traffic-control and landing systems. An example of a very different application of radar is the tracking of balloons sent aloft to gather meteorological data.

Most of the circuits, devices, and principles of operation found in radar equipment have wide application in nearly every kind of electronic equipment. Many techniques used in radar are doubly important because they appear also, for example, in television, in pulse-time communication systems, in industrial electronic devices, in equipment for measuring the height of the ionosphere, and in high-speed computing machines.

1. Radar Determination of Range. A number of different methods have been used to determine the range of objects that reflect radio waves. Two transmitter-receiver sets separated a considerable distance can be used for triangulation, as in optical ranging, but the method is cumbersome, and the required angular accuracy is hard to obtain with radio waves. Better methods time the interval required for the radio wave to make the trip from radar set to object and back. Because the speed of propagation of radio waves is known with great accuracy, range can be determined as accurately as the time interval can be measured. In one scheme, the frequency of a transmitter is changed at a constant rate and

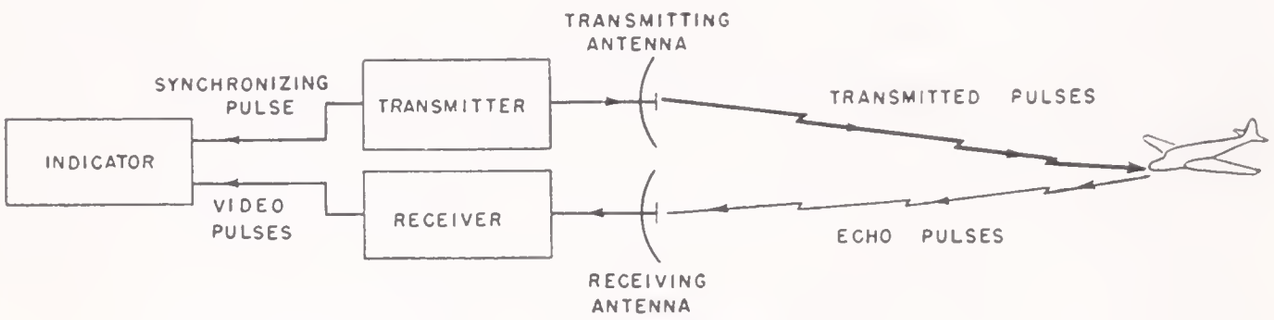


FIG. 1. Elementary radar set.

the distance to the reflecting object determined from the frequency difference between the echo and the wave being transmitted at the time the echo returns.[1] By far the most important range-measuring method, which probably had its origin in the measurements of height of the ionosphere,[2] employs a transmitter that operates in bursts of very short duration and measures range in terms of the time interval between transmission of a *pulse* and reception of an *echo pulse*.[3] So important is *pulsed*

[1] I. WOLFF and D. G. C. LUCK, "Principles of Frequency-modulated Radar," *RCA Rev.*, **9** (March, 1948), 50–75.

[2] For a historical background of radar and bibliography, see HENRY GUERLAC, "The Radio Background of Radar," *J.F.I.*, **250** (October, 1950), 285–308.

[3] The pulse-echo principle is employed in many other devices for the measurement of range. One important example is *sonar* equipment, which utilizes underwater pulses of sound. See R. J. EVANS, "Echo-ranging Sonar," *Electronics*, **19** (August, 1946), 88–93.

radar that it is usually implied by the word *radar*, and it alone is the subject of this book.

The basic principle of pulsed radar is illustrated by the block diagram of an elementary radar set in Fig. 1. This set comprises a *transmitter* and *antenna* arranged to send out radio waves in a series of short pulses, an antenna and *receiver* for detecting the echo pulses that return, and an *indicator*. The indicator includes a cathode-ray tube, and distances on the screen of this tube are proportional to the time interval between transmitted and echo pulses and thus to the range of the object. A photograph of one of the simplest radar sets appears in Fig. 2.

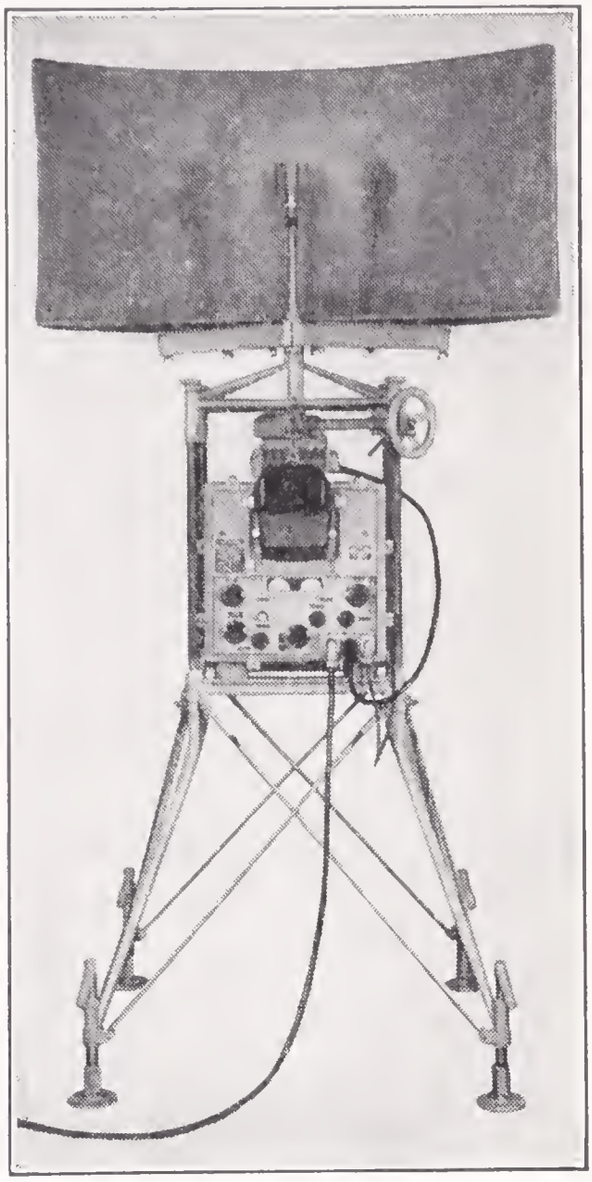


FIG. 2. Example of a low-power lightweight radar set (the SN). The SN is an early-model shipboard set and was used for search purposes in emergencies. Like most radar sets, the SN uses a single antenna for both transmission and reception. The unit beneath the antenna contains transmitter, receiver, and indicator circuits. The radio frequency is approximately 2,500 Mcps, the pulse duration is 1 μs, the repetition frequency is 1,000 cps, and the peak power is about 3 kw. (*Courtesy of General Electric Company.*)

The transmitter of Fig. 1 supplies radio-frequency energy in pulses of high power but brief duration to the transmitting antenna. The waveform of the transmitter output voltage is therefore as indicated in Fig. 3*A*. During a short time interval, the transmitter produces a sinusoidal voltage wave of extremely high frequency —say 3,000 Mcps (3,000,000,000 cycles per second)—indicated by the closely spaced lines in Fig. 3*A*. During the remainder of the time, the transmitter is quiescent. The time of operation of the transmitter, called the *pulse duration* T, varies from one radar set to another, but a typical value for it is 1 μs (one microsecond, or one-*millionth* of a second). The total time required for the completion of a cycle of events in the radar system—the pulse duration plus the time between pulses, during which echoes may be received—is called the *repetition period* T_r. The reciprocal of T_r is the *repetition frequency* f_r, or number of repetitions of the cycle of events that occurs in one second. A typical value of T_r is 2,000 μs, and the corresponding value of f_r is 500 cps. Note that the transmitted

pulse lasts for only a tiny fraction of the repetition period—so small a fraction, indeed, that it is impossible to draw the pulse to correct time scale in Fig. 3*A*. The ratio of pulse duration to repetition period is called the *duty ratio* of the transmitter. For the typical values of T and T_r assumed, the duty ratio is 0.0005.

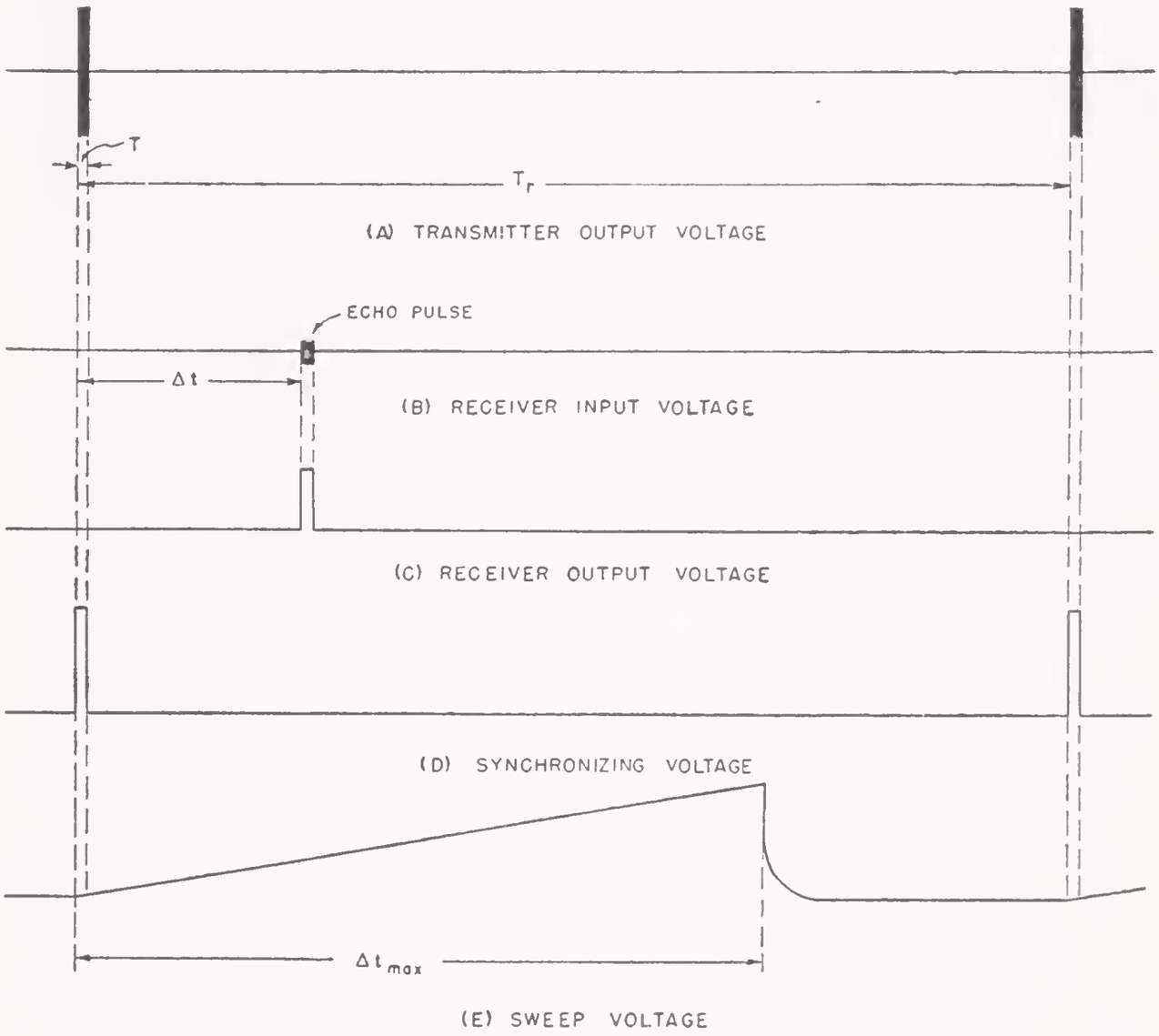


FIG. 3. Waveforms showing time relations in elementary radar set.

The transmitter is required to produce a tremendous burst of power during its brief period of operation. The energy of the transmitted pulse must be large in order that the minute part of it returned in the echo may be detectable.[1] Because the pulse is very short, extremely large transmitter power is required to produce a pulse of sufficient energy. A typical value of power produced during the pulse, called the *peak power* of the transmitter, is 100 kw. Because the duty ratio is small, the *average power* output of the transmitter is not large, despite the high peak power.

[1] The discussion in Art. 4 indicates that the ease of detecting the presence of echo pulses of different lengths depends on the energies of the pulses rather than on the peak powers of the pulses.

Because

$$\frac{\text{Average power}}{\text{Peak power}} = \frac{T}{T_r} = \text{duty ratio} \tag{1}$$

the average power corresponding to the values assumed is 0.0005×100 kw, or 50 watts.

Radio waves that travel from the transmitting antenna to the reflecting object and back to the radar set generate a radio-frequency voltage in the receiving antenna. If only a single small object reflects the waves, therefore, the waveform of receiver input voltage is as in Fig. 3*B*. The time required for the two-way trip of the waves is called Δt. Voltage amplitudes in Figs. 3*A* and *B* should not be compared; the echo pulse would be completely invisible if the two figures were drawn to the same scale. The transmitted-pulse power is of the order of 100 kw, but the peak power of the echo pulse may be less than 1 $\mu\mu$w (0.000,000,000,001 watt). The function of the receiver is to amplify this tiny pulse to usable size and to detect it, that is, to produce a direct voltage that varies in magnitude in accordance with the envelope of the radio-frequency echo pulse. The receiver output is therefore a pulse of direct voltage, called a *video pulse*, as indicated in Fig. 3*C*.

The time interval Δt in Fig. 3 is equal to the time required for a radio wave to travel from the radar set to the object detected and back. If r denotes the distance from set to object and c the speed of light (which, for practical purposes, equals the speed of propagation of radio waves in air), the time required for the two-way trip is $\Delta t = 2r/c$ and the range is

$$r = \frac{c}{2}\Delta t \tag{2}$$

This fixed proportional relation between range and the delay time of the echo pulse is the basis of range measurement by radar.

Conversions between time intervals and ranges are so frequently required in the study of radar that a list of values of $2/c$ and $c/2$ for several convenient units of length is useful. From the value of c, 2.998×10^8 m per sec, or 186,000 miles per sec, the constants in Table 1 are calculated.[1] Persons working constantly with radar find it convenient to remember the italicized numbers in this table.

Though long relative to the pulse duration, the time interval Δt is much too short to be measured by mechanical means. If, for example, the

[1] The speed of radio waves in air is about 2.985×10^8 m per sec, slightly less than the speed of light in vacuum. See R. C. COLWELL, H. ATWOOD, J. E. BAILEY, and C. O. MARSH, "The Velocity of Radio Waves over Short Paths," *Proc. I.R.E.*, **30** (1942), 129–131. Nevertheless, the figures of Table 1 based on the speed in vacuum are the ones commonly employed. Actual determination of range by radar always depends on a direct calibration of the range circuits of the set, so that knowledge of the exact speed of the waves is not important.

TABLE 1. RANGE-TIME CONVERSION FACTORS

$c/2$		$2/c$	
0.0810	nautical miles per μs	*12.36*	μs per nautical mile
0.0931	statute miles per μs	*10.74*	μs per statute miles
0.164	thousand yd per μs	*6.10*	μs per thousand yd
492	ft per μs	0.002035	μs per ft

range of an object is 50 statute miles, Δt is only $50 \times 10.74 = 537$ millionths of a second. To measure this short time, the radar indicator employs electronic circuits, and the results of the measurement are presented on the screen of a cathode-ray tube. In the simplest scheme, the receiver output voltage is applied to one pair of deflecting plates in the cathode-ray tube to cause a vertical motion of the spot on the screen, and a saw-tooth *sweep voltage* is applied to the other pair of plates to cause horizontal motion of the spot. The appearance of the indicator tube is illustrated in Fig. 4. The approximate range of the object from which each echo signal is obtained can be read from a scale engraved on a transparent screen placed over the face of the tube.

Voltages required to produce the presentation of Fig. 4 are shown in Figs. 3*D* and *E*. The indicator receives a *synchronizing voltage* from the

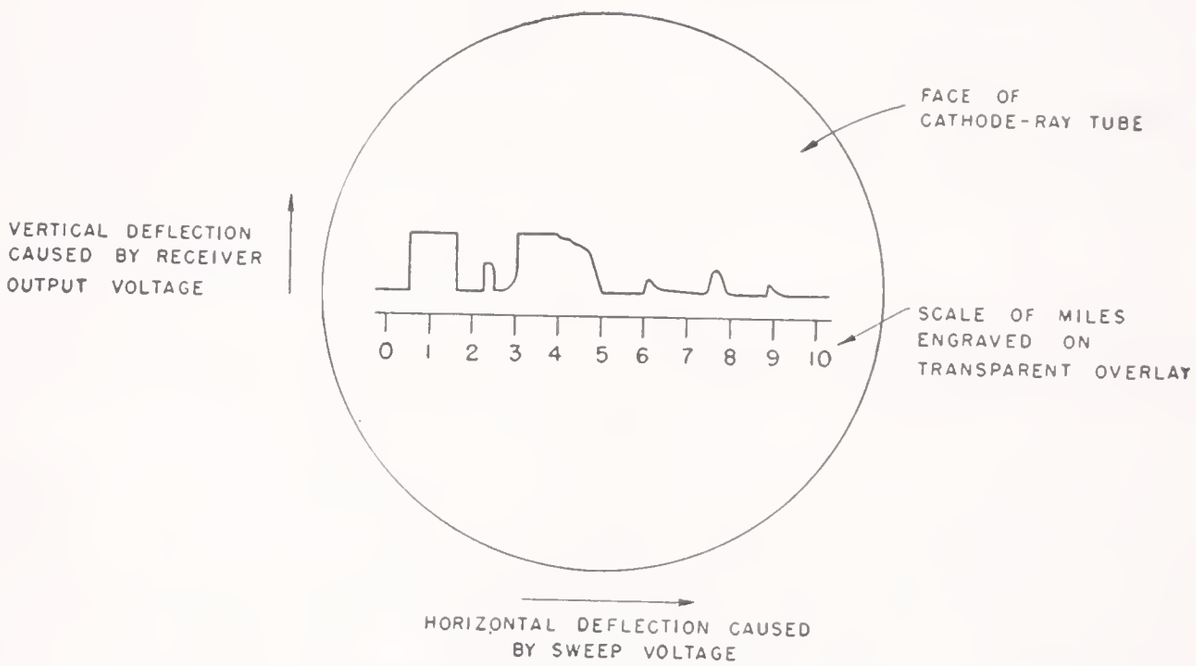

FACE OF CATHODE-RAY TUBE
VERTICAL DEFLECTION CAUSED BY RECEIVER OUTPUT VOLTAGE
0 1 2 3 4 5 6 7 8 9 10
SCALE OF MILES ENGRAVED ON TRANSPARENT OVERLAY
HORIZONTAL DEFLECTION CAUSED BY SWEEP VOLTAGE

FIG. 4. General appearance of simple indicator screen.

transmitter. This voltage is a video pulse occurring at the same instant as the radio-frequency transmitted pulse. The leading edge of the synchronizing pulse marks the beginning of the time interval to be measured, and the leading edge of the echo pulse marks the end of the interval. The time at which the saw-tooth sweep voltage (Fig. 3*E*) begins is, therefore, controlled by the leading edge of the synchronizing pulse. The value of the sweep voltage at the starting instant is chosen so that the indicator spot is deflected nearly to the left-hand edge of the

indicator screen. After the first instant, the sweep voltage rises at a constant rate and the spot moves across the screen at constant speed. The rate of rise of sweep voltage is adjusted so that the spot is nearly at the right-hand edge of the screen at the end of a time interval Δt_{max}, corresponding to the maximum range at which objects are to be located. After the end of the sweep interval, the sweep voltage returns to its initial value and the spot moves back to the left-hand side of the screen, in preparation for the next cycle of operation. The pattern of Fig. 4 is produced during the sweep interval, because vertical motion resulting from the receiver output voltage is superimposed on the left-to-right sweep motion, resulting in *pips* corresponding to the echoes from various objects. Because the sweep speed is constant, horizontal distances on the indicator screen are proportional to delay times of the echoes and therefore to range. Thus the range scale engraved on the overlay is linear.

The pips in Fig. 4 might be expected to indicate, by their lengths along the range scale, the extent in range of the objects that cause them. If the transmitted pulse were of negligible duration, this expectation would be realized. However, because the transmitted-pulse duration is greater than zero, the apparent extent of each object is increased by the range equivalent of the pulse duration—even the smallest object cannot return a pulse of shorter duration than the transmitted pulse. Thus, with a 1-μs pulse, an object extending 1,000 ft in range causes an echo pip covering about 1,500 ft on the range scale, and if two objects are separated by less than 500 ft, their echo pulses run together and the radar set cannot show them separately. The ability to separate objects that are close together is called *resolution*. The *range resolution* of a radar set is determined by the duration of the transmitted pulse. Need for good resolution is the principal reason why such short transmitted pulses are used in radar.

2. Radar Direction Indication. The description of an elementary radar set in Art. 1 explains a method of measuring *range* but does not discuss the equally important measurement of *direction angles*. Measurement of angles by radar depends upon the use of highly *directional* antennas—transmitting antennas that send out radio waves in a narrow beam and receiving antennas that respond most strongly to waves arriving from one particular direction.

If the transmitting antenna is directional, azimuth and elevation angles of an object may be determined by orienting the antenna so that maximum radio-wave illumination is received by the object—just as a searchlight may be turned until the brightest part of its beam strikes the object. Maximum radio-wave illumination is indicated by the reception of the strongest echo pulses, and the object direction can be determined

from the corresponding antenna orientation. If the receiving antenna is directional, it may be turned until the largest echo pulse is received from the object. Azimuth and elevation angles are then indicated by the receiving-antenna orientation.

If both antennas are directional, and if the two antennas are turned as a unit so that the direction of maximum radiation is always the same as the direction of best reception, then the orientation of the two-antenna unit becomes an indicator of direction. Use of two directional antennas in this way improves the accuracy of angle measurement, because a slight shift of the antennas from the correct position has a double effect in reducing the amplitude of the pulse observed; the antennas can therefore be positioned more accurately.

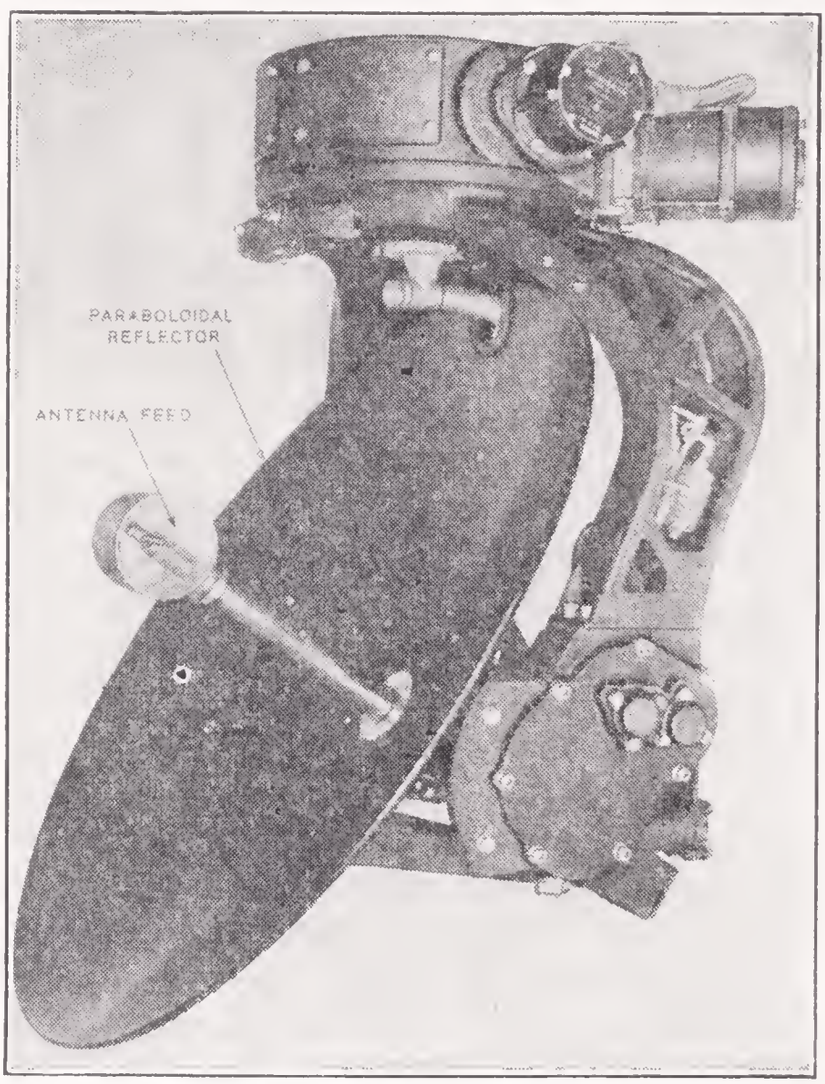


FIG. 5. Example of a 10-cm radar antenna. The reflector is 29 in. in diameter. The beam angle is 10 deg. (*Courtesy of Bell Telephone Laboratories.*)

Photographs of two directional radar antennas appear in Figs. 5 and 6. The antenna of Fig. 5 resembles somewhat in appearance and principle of operation the ordinary reflector-type searchlight. The *antenna feed* in Fig. 5 is a relatively nondirectional source of radio waves and corresponds to the arc in a searchlight. The reflector is a paraboloid of revolution

and, like the reflector of a searchlight, focuses the waves into a narrow beam. Antennas of this type are widely used in radar, especially at frequencies above 1,000 Mcps. At lower frequencies, antennas of the type illustrated in Fig. 6 are sometimes used. These antennas resemble antenna arrays employed at still lower frequencies in radio communication.

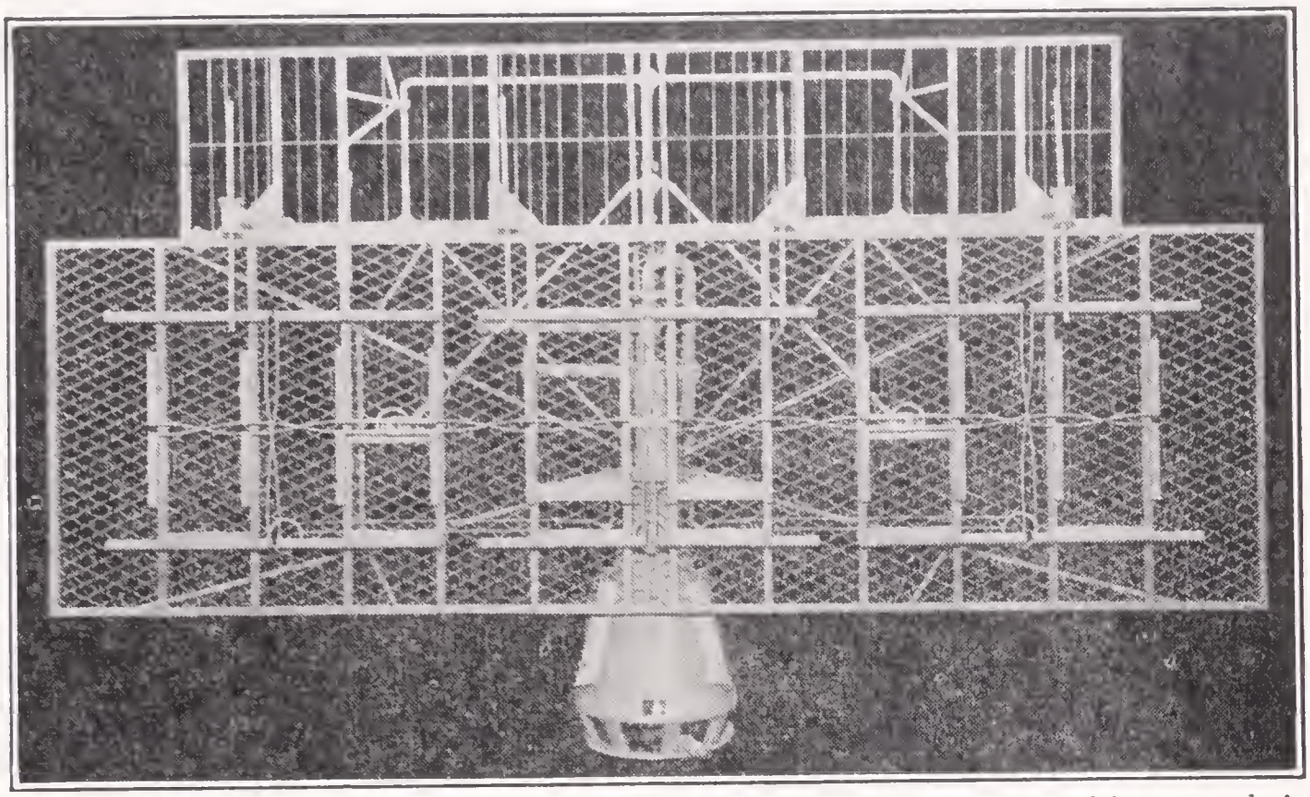


FIG. 6. Example of an antenna for a 200-Mcps radar set. The horizontal beam angle is 20 deg, and the vertical beam angle is 60 deg. (*Courtesy of General Electric Company.*)

Despite the resemblance of the antenna of Fig. 5 to a searchlight, radio waves, because of their longer wavelength, are much more difficult to confine to narrow beams than are light waves. In a searchlight, the narrowness of the beam is limited almost entirely by the size of the source of light and the geometrical perfection of the reflector surface. Even if a radar antenna could be built with a feed that operated as a point source and a perfectly paraboloidal reflector, the beam of radio waves produced might still be very wide. Because of diffraction effects, *a narrow beam can be produced only if the dimensions of the antenna are large relative to the length of the radio waves.*[1] Therefore, to obtain narrow beams and accurate direction determination, either the antenna must be very large or the wavelength of the radar set must be very short.

In order to obtain the greatest accuracy in determining direction angles, radar sets commonly employ directional antennas for both transmission and reception. To avoid the serious mechanical problems associated with using two large antennas and rotating them so that axes remain

[1] This principle is stated more accurately and explained more fully in Art. 13, Chap. XIII.

accurately aligned, means have been developed for employing only one antenna and using it for both transmission and reception. The directivity of the antenna is thus used twice, and the axes of maximum radiation and maximum sensitivity of reception are automatically aligned, provided the antenna is not moved appreciably during the time interval between transmission of a pulse and reception of the echoes (an interval usually less than 1/500 sec).

The use of a single antenna for both transmission and reception in nearly all radar sets is made possible by a *transmit-receive device* (abbreviated t-r device) connected as in Fig. 7. The t-r device functions as a

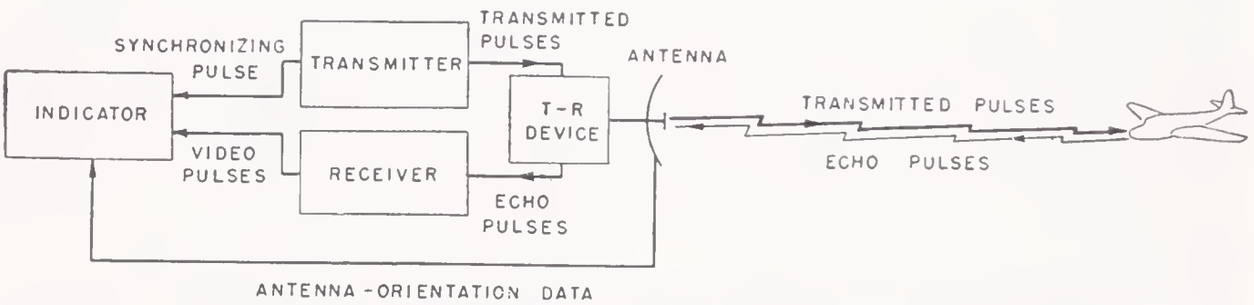


FIG. 7. Radar set employing single antenna for both transmission and reception.

switch (see Arts. 10 to 14, Chap. XII) to connect the antenna to the transmitter during the pulse and to the receiver during the remainder of the repetition period. The t-r device is also called a *duplexer* and a *receiver-protective device.* The last name is applied because the device must protect the sensitive input circuit of the receiver by limiting to a small value the portion of the transmitted power that reaches the receiver. Figure 7 differs from the elementary block diagram of Fig. 1 not only because a single antenna and a t-r device are indicated but also because a connection from antenna to indicator is shown. This connection provides antenna-orientation data at the indicator, so that the operator may know the direction as well as the range of objects corresponding to the pulses he sees.

The narrowness of the antenna beam is important because it determines the accuracy with which the radar set can measure azimuth and elevation angles. It is important also because it determines the *angular resolution* of the radar set, exactly as pulse duration determines the range resolution. The apparent angular extent of any object is increased by the width of the antenna beam, and objects separated by less than the width of the beam appear to the radar operator as a single object.

Figure 8 is a pictorial representation of the radiation from a directional antenna. As the waves travel outward from the antenna, wavefronts form spherical surfaces that expand radially. The total power transmitted by the antenna is distributed over the spherical surface of a wave-

front, but because the antenna is directional, the distribution of power over the surface is not uniform. The *power density* (power per unit area of the wavefront) is great at points within the beam, as A, and small at points outside the beam, as B. The edge of the beam is not sharp; the power density decreases gradually from a maximum at point A on the beam axis to zero at a point such as C. Beyond C, the power density varies between zero and small nonzero values. The region of main radiation between C and C' is called the *major lobe*, and regions of small radiation are called *minor lobes*. Points of no radiation, like C, are called *nulls*.

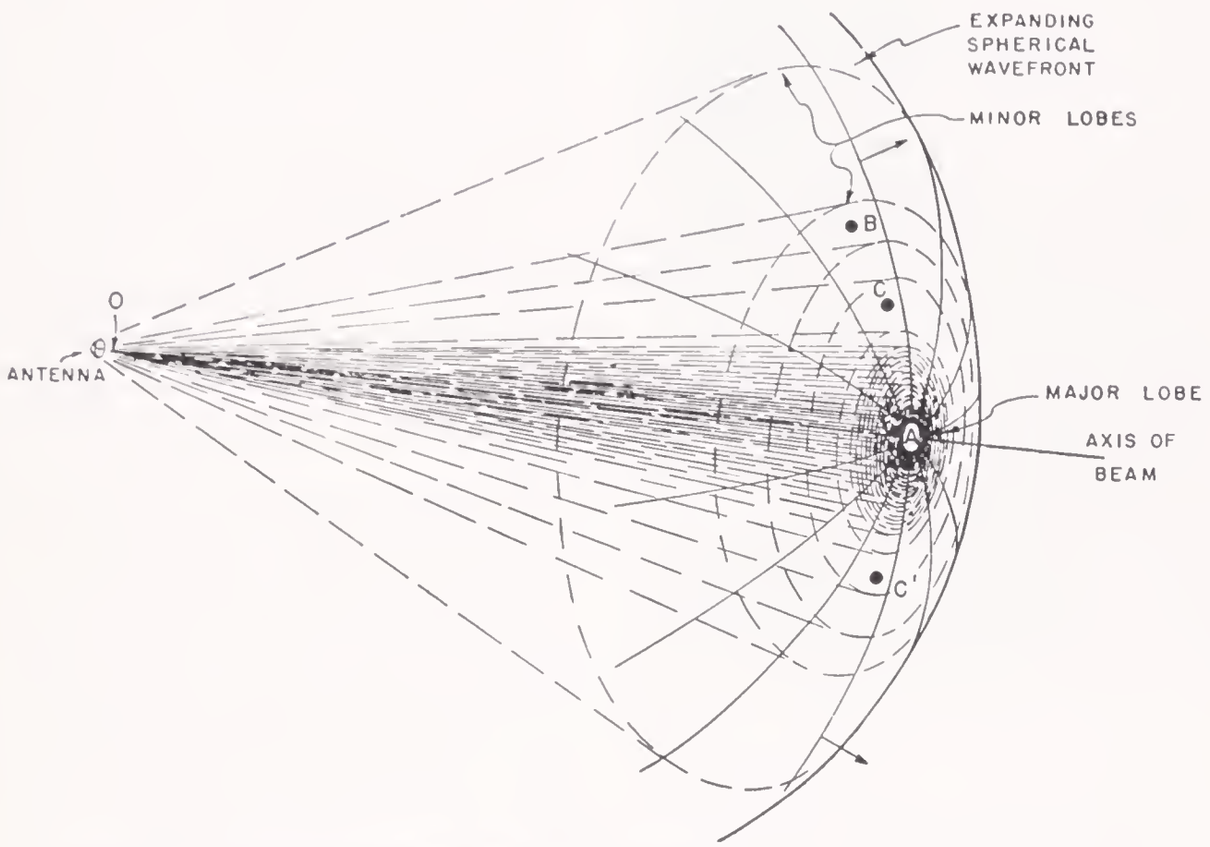


FIG. 8. Pictorial representation of radiation from directional antenna.

The width of the beam, the major lobe in Fig. 8, is usually specified by the *beam angle between half-power points*. This is the angle between lines on opposite sides of the beam axis OA along which the power density is half as great as it is on the axis. The beam angle serves as a measure of the angular accuracy and resolution of a radar set. The beam in Fig. 8 has a circular cross section, but beams of other cross-sectional shapes are often used. For example, by using the truncated-paraboloidal reflector of Fig. 9, having a horizontal dimension *greater* than the vertical dimension, the horizontal beam angle is made *smaller* than the vertical beam angle. Such a reflector may be used on a shipboard radar set to search the surface of the sea. The small horizontal beam angle provides good azimuth resolution, while the large vertical angle allows surface search to continue despite rolling and pitching of the ship.

3. Data Presentation. When both direction and range information is required, methods of data presentation different from the simple one described in Art. 1 are often desirable. Many different data-presentation methods have been developed and have been given type designations. The presentation described in Art. 1, in which echo strength is indicated by vertical displacement of the luminous spot and range by horizontal displacement, is called *Type A* data presentation. Other important presentations are *Type B*, in which vertical displacement indicates range and horizontal displacement indicates azimuth angle, and *Type P*, in which *radial* displacement of the spot from the center of the screen indicates range and the direction of the radial displacement indicates azimuth angle. Type P presentation is usually called *plan-position indication* (abbreviated PPI). When Type A presentation is used, the electron beam of the cathode-ray tube is said to be *deflection modulated*, because video pulses from the receiver cause deflection of the beam. With Type B presentation and plan-position indication, range and azimuth signals control the deflection of the electron beam and video pulses from the receiver control the intensity of the luminous spot. The electron beam is said to be *intensity modulated*.

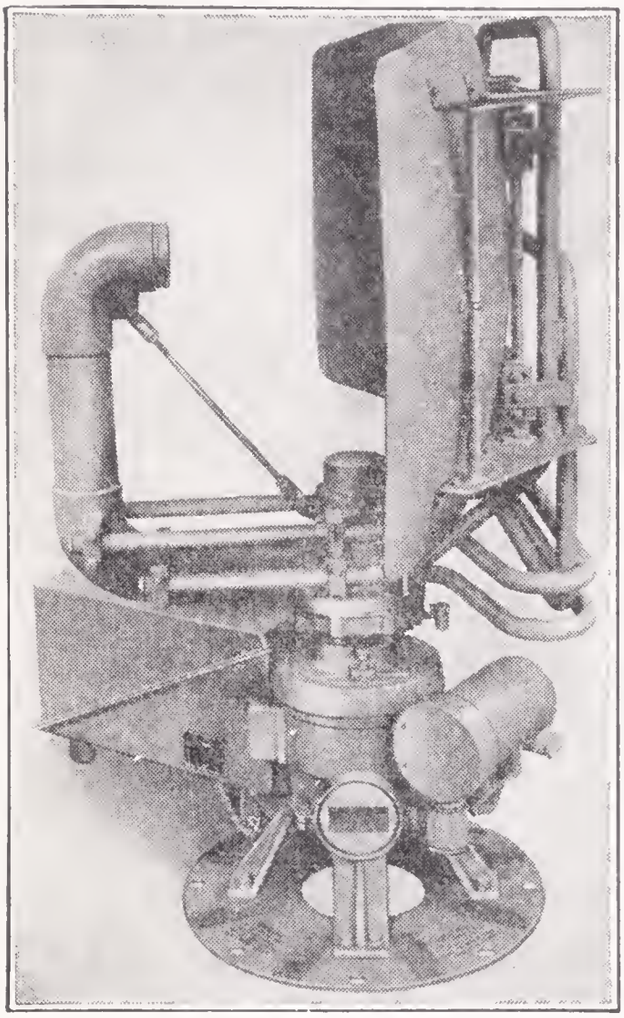
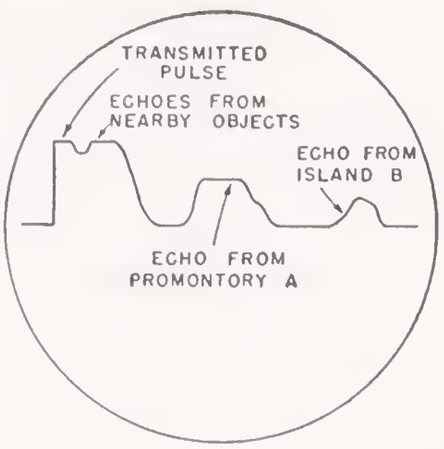


FIG. 9. A truncated-paraboloidal antenna for 10-cm operation. The reflector is a 20-in. sector of a 42-in. paraboloid. Because the horizontal dimension of the antenna is greater than the vertical dimension, the beam is narrower in the horizontal plane than in the vertical plane. The horizontal beam angle is 6 deg, and the vertical beam angle is 12 deg. (*Courtesy of Bell Telephone Laboratories.*)

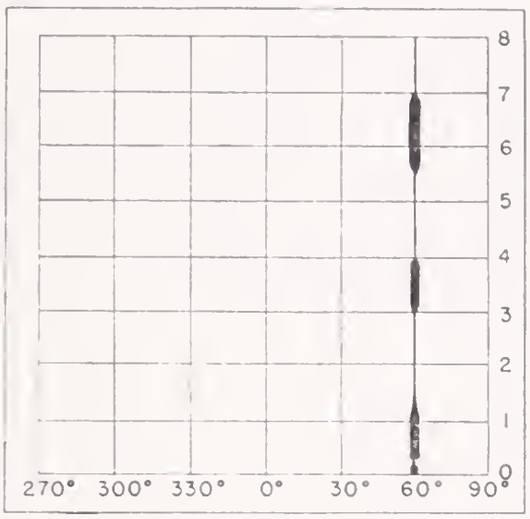
Consider a radar set located on a coast line at the position indicated by the center point in Fig. 10*A*. Assume that the antenna of the set radiates waves in a narrow horizontal beam. The antenna can be rotated about a vertical axis, and thus the radar set can *scan* the surface of the surrounding land and sea. The displays of information obtained with Type A, Type B, and plan-position presentations are to be examined.

If Type A presentation is employed, a different pattern is obtained on the indicator screen for each position of the antenna. If the antenna

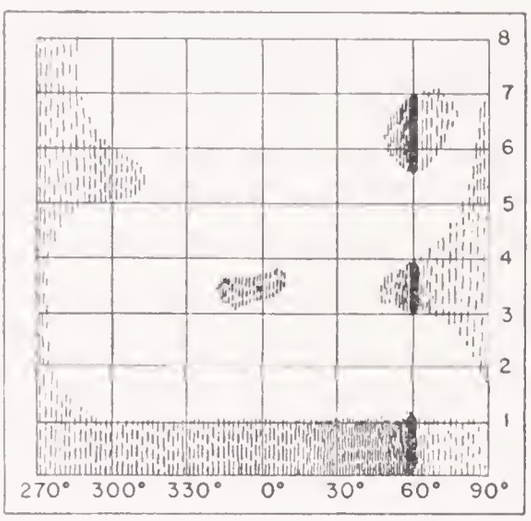
(A) LOCATION OF RADAR SET



(B) TYPE A PRESENTATION FOR AZIMUTH ANGLE OF 60°

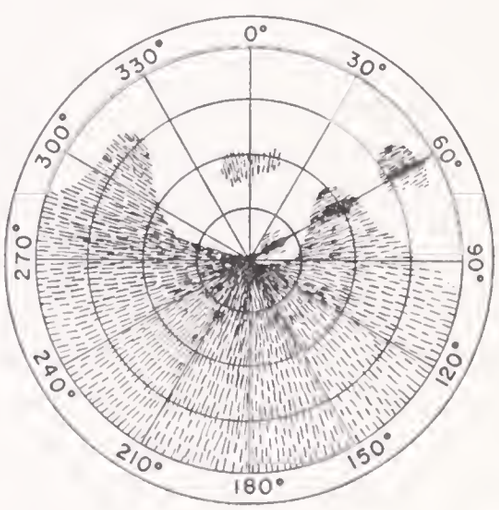


(C) TYPE B PRESENTATION — ANTENNA STATIONARY AT 60° AZIMUTH ANGLE



(D) TYPE B PRESENTATION — ANTENNA ROTATING

(E) APPEARANCE OF PPI — ANTENNA STATIONARY AT 60°



(F) APPEARANCE OF PPI — ANTENNA ROTATING

FIG. 10. Relation of Type A, Type B, and plan-position presentations to map of region surrounding radar set.

beam lies along the 60-deg azimuth line in Fig. 10*A*, the presentation of Fig. 10*B* is obtained. Reflections from promontory *A* and island *B* cause deflections of the trace on the cathode-ray tube as indicated. If the operator records the indicator pattern obtained for each of many positions of the antenna, he can construct from these patterns a rough map of his surroundings. Such a process, however, would be tedious. Other types of data presentation offer quicker, more direct means of visualizing the topography of the region.

The pulses at the beginning of the sweep in Fig. 10*B* should be noted. Similar pulses appear in most radar displays. At the very beginning is a signal caused by the transmitted pulse. The transmit-receive device is not perfect and allows some of the transmitted-pulse power to reach the receiver input circuit. Though only a tiny fraction of the transmitted power, this power is large relative to echo-pulse power and causes a large deflection on the indicator screen. Following the transmitted pulse, a clutter of echoes appears. One cause of these echoes is the radiation in minor lobes of the antenna pattern. Large near-by objects (for example, the land or sea directly under the radar antenna) cause strong echo pulses because of their large size and short ranges—despite the small amount of radiation in their direction. Clutter at somewhat greater ranges is obtained because waves in the sea and small irregularities in relatively level land surfaces return energy to the radar set. Thus *sea return* is obtained from a point such as *C* in Fig. 10*A*. The short range to *C* compensates for the relatively poor ability of the sea surface to return energy to the radar set.

If the data-presentation method is changed from Type A to Type B, the horizontal range sweep of Fig. 10*B* is changed to the vertical one of Fig. 10*C* and at the same time vertical deflection of the trace by echo signals is replaced by a modulation of the intensity of the trace. The appearance of the Type B trace when the antenna is stationary at an azimuth angle of 60 deg is suggested by Fig. 10*C*. The trace consists of a faint vertical line, portions of which are made bright by the echo signals. The range and azimuth coordinate lines and markings in Fig. 10*C* represent lines and numbers engraved on a transparent cover for the indicator screen. Figures 10*B* and *C* are drawn for the same azimuth angle and may be compared. The trace in Fig. 10*C* is shown at the 60-deg position instead of at the center of the screen because in Type B presentation the trace is moved sideways when the antenna is turned, so that the horizontal position of the trace indicates the azimuth angle of the antenna.

Type B presentation is most useful if the antenna scans continuously. Suppose that the radar set of Fig. 10*A* is intended to view only the 180-deg region facing the sea and that the antenna rotates clockwise at a constant speed. As the antenna turns from the 270-deg position through

0 deg to the 90-deg position, the bright line in Fig. 10*C* moves from left to right across the screen and the pattern of bright spots on the line changes as it moves. The indicator screens used for Type B presentation have very long *persistence*, or afterglow. Therefore, fading smears of light are left behind the bright line as it moves, and the presentation produced has an appearance suggested by Fig. 10*D*. The signals obtained while the antenna rotates through the second half circle, from 90 to 270 deg, are not of interest; therefore the electron beam in the cathode-ray tube is cut off during this interval, and nothing appears on the screen except the fading picture left by the preceding trace.

The glowing spots of light represented by Fig. 10*D* resemble in a remote way a map of the region viewed by the radar set. The most important difference between a Type B display and a maplike picture of the region is the distortion caused by representing *angles* of azimuth by horizontal *displacements*. To correct the distortion, the vertical coordinate lines in Fig. 10*D* would have to be moved without change of length in such a way that their bottom points coincide and their top points swing outward until the 90- and 270-deg lines are horizontal. A corresponding alteration of the light-spot pattern would make the display of Fig. 10*D* resemble the top half of Fig. 10*A*. The distortion in Type B presentation can be made small if the indicator is designed to cover only a small sector of the region about the radar set. For example, if the full width of the indicator screen corresponds to angles from 0 to 30 deg and the full length to ranges from 6 to 8 miles, the distortion is not very noticeable.

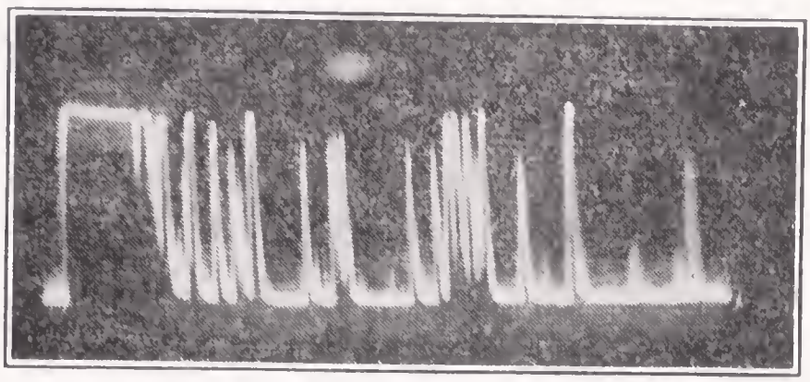
Several other effects prevent Type B presentation from producing a perfect maplike picture. One is the clutter obtained at short ranges, which may mask objects in this region. Another is the failure to indicate distant objects that lie in the *radio shadow* of near-by objects. Thus, in Fig. 10*A*, if promontory *A* is large and high relative to island *B*, radio waves may be prevented from reaching the island, and the radar indication cannot then show the existence of the island. Still another effect is the *resolution* of the set. Objects are spread out in range because of the pulse duration and in azimuth because of the azimuth beam angle. A group of small, closely spaced objects appears on the Type B indicator as a single large object. Because most radar sets employ pulses with very steep sides, the range separation required to resolve two objects is rather definite. The angular separation, however, is less definite because the edges of the antenna beam are not sharp. If the gain of the radar receiver is increased, the echo caused by a weak part of the antenna beam may become perceptible, and thus the apparent spread of the object in azimuth may be increased. Consequently, best resolution is obtained with the minimum gain setting that provides an observable signal. Echo intensity falls off rapidly with range. Therefore a single setting of gain

is not suitable to all ranges. If the gain is adjusted so that distant features are visible, the resolution of near-by objects is poor.

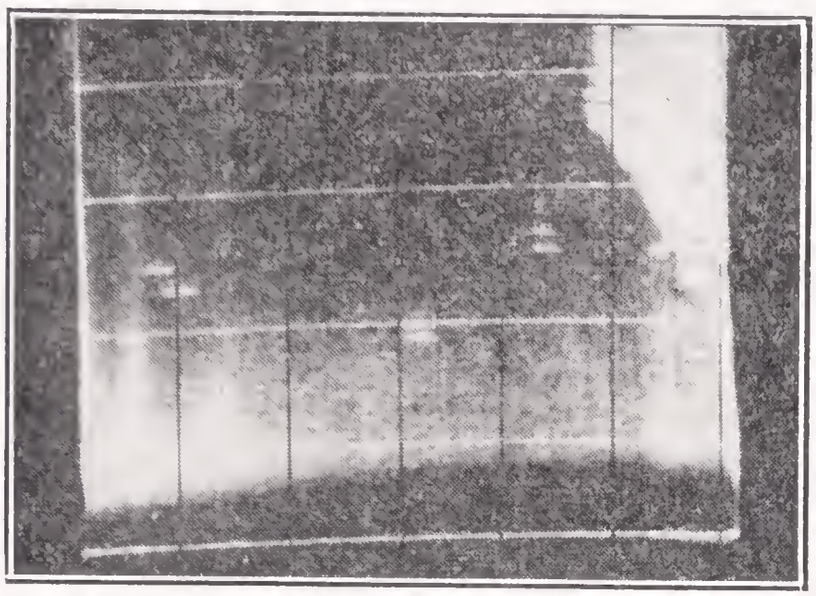
The distortion of coordinates inherent in Type B presentation is avoided in plan-position indication, but the other defects enumerated are common to both presentations. The Type P screen often has a transparent overlay carrying range circles and azimuth lines as in Fig. 10*E*. The range trace appears as a line with bright spots extending radially outward from the center of the screen. For the antenna stationary in the 60-deg position, the appearance of the screen is as indicated by Fig. 10*E*. This figure may be compared with the Type B presentation of Fig. 10*C*—the only difference is the position of the range trace. Turning the antenna causes the plan-position range trace to turn about the center of the screen in such a way that the angle of the trace always is the same as the azimuth angle of the antenna. Therefore, when the antenna rotates, the presentation suggested by Fig. 10*F* is obtained. The bright line rotates around the screen, leaving fading smears of light behind it. The glowing pattern produced is a rough map of the region around the radar set (compare Figs. 10*A* and *F*).

4. Discernment of Signals and Noise. The diagrams of Fig. 10 showing Type A, Type B, and Type P presentations are idealized in one very important respect: No indication of *noise* appears in them. Noise is the name given to very small randomly fluctuating voltages that unavoidably are present *in the input circuits of the receiver.* (The causes of noise are listed and explained in Arts. 13 to 17, Chap. VI.) The word *noise* is a carry-over from sound-communication terminology. Noise voltages in sound equipment produce actual noise—a hissing sound—in the loudspeaker output. In radar, noise voltages result in erratic, random deflections of the electron beam of a Type A indicator or correspondingly erratic fluctuations of intensity in Type B and plan-position presentations. The appearance of noise and the way in which noise can obscure weak signals are indicated by the photographs of radar-indicator presentations in Fig. 11.

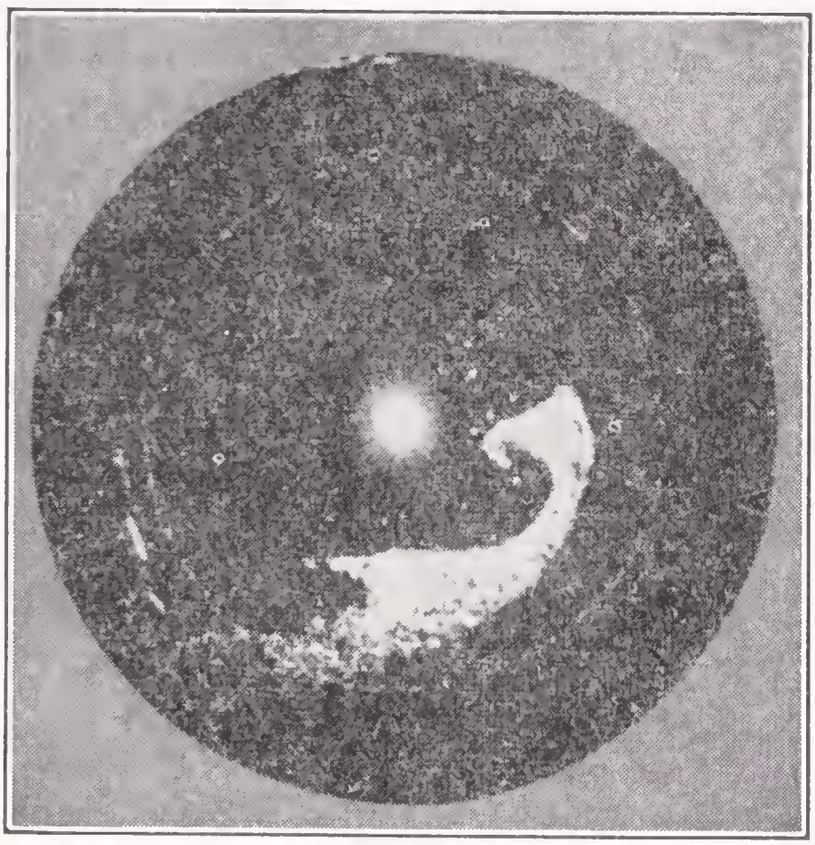
Noise voltages are entirely unwanted but cannot be avoided in physical circuits. Were it not for noise, the maximum range at which an object is detectable by radar could be extended almost indefinitely. Objects at great range return exceedingly weak echoes, but without noise almost any signal could be amplified to a usable level if enough stages were added to the receiver. Because of noise, however, the limit of sensitivity is reached when the signal level falls below the noise to such an extent that the signals are obscured. An increase of amplification is then of no avail, because signals and noise are amplified together. Radar receivers are always designed with enough stages so that large-amplitude noise voltages



(*A*) Type A



(*B*) Type B



(*C*) Type P

FIG. 11. Photographs of radar-indicator presentations showing echo signals and noise.

appear in the output when full gain is used, and the limit of sensitivity is therefore set by the noise.

Because noise does limit receiver sensitivity, *the input circuits of the receiver are the most critical parts of the entire radar system.* To make possible the detection of a given object at maximum range, the input circuits must utilize the small echo signal as effectively as possible and at the same time must combine with the signal the least possible amount of noise. That is, the highest obtainable *signal-to-noise ratio* is desired in order that the signal at the indicator may be discernible. Subsequent receiver stages are not nearly so critical, because they amplify both signal and noise at relatively high levels and the noise they introduce is negligible compared with that already present. In the input circuits, the loss of one micromicrowatt of echo signal or the addition of an equally small noise power can reduce the maximum range of the radar set as seriously as would the loss of many kilowatts of transmitter power.

The noise limitation of receiver sensitivity is part of the reason for the statement in Art. 1 that the ease of detecting the presence of an echo pulse depends upon the energy of the pulse rather than upon its peak power. It might seem that peak power is more important, because by halving the duration of a pulse, the peak power could be doubled without changing the energy, and the greater amplitude pulse thus obtained should stand out more clearly from the noise. The error in this argument involves the *bandwidth* of the receiver. In order to amplify pulse signals, the receiver must be responsive to a band of frequencies centered about the nominal frequency of the echo pulse,[1] and the width of the frequency band required is inversely proportional to the pulse duration. Thus, halving the echo-pulse duration requires that the receiver bandwidth be doubled. Unfortunately, noise power is proportional to receiver bandwidth, and thus noise power also is doubled. Therefore, when the duration of a constant-energy pulse is halved, not only is the power of the pulse doubled but also the power of the noise with which the pulse must be compared is doubled, and therefore no change in the discernibility of the signal is obtained. The true sensitivity of a radar receiver is measured by the *energy* content of the smallest energy input pulse that can produce an output signal discernible against the background of noise.

5. Radar Range Equation. Very important in the application of a radar set is its *maximum range,* that is, the range beyond which an object cannot be detected because echo signals are obscured by noise. The range attainable depends on the reflecting properties of the object as well as on the characteristics of the radar set. Not only do some objects return more of an incident radio wave than do others (an island would produce a much larger echo than a buoy, for example, and would be

[1] For discussion of bandwidth requirements, see Art. 18, Chap. VI.

detectable at greater range), but also the reflection from a given object depends upon the orientation of the object. For this reason, radar signals from ships and aircraft appear to twinkle (in an indicator display using intensity modulation), because they change amplitude from moment to moment with the rolling and pitching of the craft. Maximum range is affected also by the location of the radar set and object relative to the surface of the earth and by atmospheric conditions. These factors influence the propagation of radio waves between radar set and object (see Chap. XIV).

In this article the relation of maximum range to radar-set characteristics is to be considered. Propagation effects are therefore eliminated from the discussion by assuming radar set and object to be isolated bodies in space. The maximum range calculated on this assumption is the *free-space maximum range* of the set.

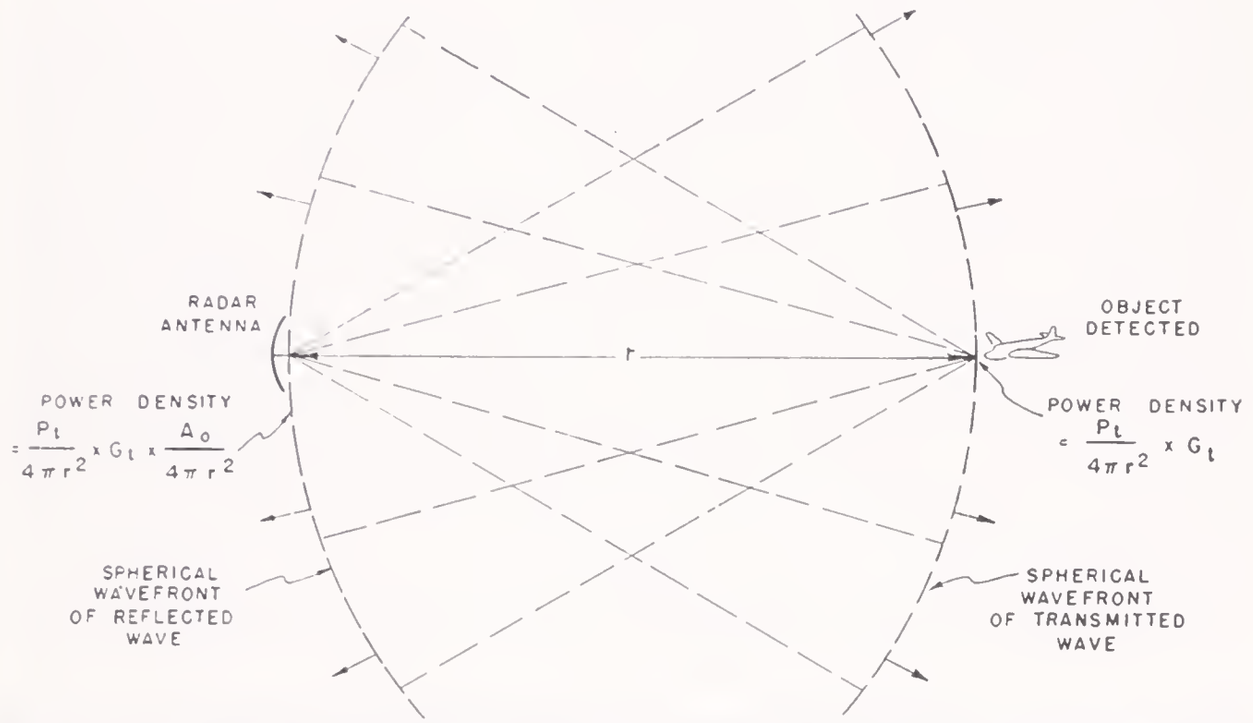


FIG. 12. Transmitted and reflected waves for radar set and object considered as isolated bodies in space.

The radar antenna and object are indicated in Fig. 12. They are separated by the range r and are considered to be isolated from all other material bodies. The antenna is positioned so that maximum radiation is directed toward the object. The peak power of the rectangular pulse of waves transmitted from the antenna is P_t. This power is spread over the spherical wavefront traveling outward from the antenna. When the wavefront reaches the object, the area of its spherical surface is $4\pi r^2$, and if the radiation were uniform in all directions, the power density at the object would be $P_t/4\pi r^2$. Because the antenna is directional, the distribution of power is not uniform. The power density is greater than

$P_t/4\pi r^2$ at the object, and less at other positions on the wavefront. Accordingly,

$$\left.\begin{array}{c}\text{Power density of the transmitted}\\ \text{wave at the object}\end{array}\right\} = \frac{P_t}{4\pi r^2} \times G_t \tag{3}$$

where G_t is a factor greater than 1 to take account of the concentration of the radiation in the direction of the object. A fictitious antenna that radiates uniformly in all directions is called an *isotropic antenna,* and G_t is called the *power gain* of the transmitting antenna relative to an isotropic antenna. The gain G_t is the ratio of the actual power density at the object to the power density that would be produced by an isotropic transmitting antenna. The appearance of r^2 in the denominator of Eq. 3 is in accordance with the *inverse-square law*—the power density of a wave radiated from a point source isolated in space is inversely proportional to the square of the distance from the source.

The transmitted wave, on reaching the object, is in part reflected. That is, some of the energy incident on the object is converted into a second wave and radiated from the object. Wavefronts of the reflected wave are expanding spheres centered at the object. Suppose, for a moment, that the object is a sphere having a perfectly conducting surface and a cross-sectional area A_o. Such a sphere can be shown to collect the power in an area A_o of the incident wave and to reradiate this power uniformly in all directions, provided the diameter of the sphere is large relative to the length of the radio waves. The reflected wavefront expanding from the spherical object has an area $4\pi r^2$ when it reaches the radar antenna, and therefore

$$\left.\begin{array}{c}\text{Power density of the reflected}\\ \text{wave at the radar antenna}\end{array}\right\} = \frac{P_t}{4\pi r^2} \times G_t \times \frac{A_o}{4\pi r^2} \tag{4}$$

Actual objects may extract more or less energy from the transmitted wave than that contained in their cross-sectional area. Furthermore they absorb some of this energy and reradiate only the remaining part, and the radiation is usually far from uniform in direction. Nevertheless the reflecting ability of any object can be described by an area A_o, called the *radar cross section* of the object, which is the cross-sectional area of a perfectly conducting sphere that would return the same power to the radar set as does the actual object. Equation 4, therefore, applies to any object if A_o is considered to stand for radar cross section. For aircraft and ships, the radar cross section changes from moment to moment as the object changes its orientation.

The power absorbed from the reflected wave by the radar antenna is proportional to the power density of the wave when it reaches the

antenna. Therefore the received power P_r is

$$P_r = \frac{P_t}{4\pi r^2} \times G_t \times \frac{A_o}{4\pi r^2} \times A_r \tag{5}$$

where A_r is the constant of proportionality. This constant has the dimensions of area and is called the *effective area of the receiving antenna.* Effective area may be thought of as an area of the wavefront of the reflected wave from which the receiving antenna obtains the power it supplies to the receiver. For some types of radar antennas, the effective area is very nearly equal to the area of the projection of the antenna on a plane perpendicular to the axis of maximum sensitivity, but for other types of antennas there is a great discrepancy between actual and effective areas. Note that effective area is defined in terms of a wave approaching from the direction of most sensitive reception.

Equation 5 is a statement of the *inverse-fourth-power law* for radar. Because the inverse-square law applies to both transmitted and reflected waves, the echo power received is inversely proportional to the fourth power of the range—echo strength falls off very rapidly as range is increased. It should be remembered that this law has been derived for a radar set and an object isolated in space. Under some circumstances it applies to an actual set operated in proximity to the earth, but under other circumstances presence of the earth makes echo strength vary with range in an entirely different manner. Especially important is the horizon effect—under normal propagation conditions a *radar horizon* exists and limits radar ranges in the same way that the optical horizon limits the range of vision (see Art. 6, Chap. XIV).

From Eq. 5 the free-space maximum range can be obtained if P_r represents the power of the weakest discernible echo pulse. Because the discernibility of a pulse depends upon its energy, the appropriate value for P_r is $W_{r_{\min}}/T$, where $W_{r_{\min}}$ is the energy of the minimum discernible pulse and T is the pulse duration. (The object detected is assumed to have small extent in range, so that the echo-pulse duration is the same as that of the transmitted pulse.) When P_r in Eq. 5 is replaced by $W_{r_{\min}}/T$, r becomes the maximum range $r_{\max}$, and the resulting equation solved for $r_{\max}$ is the *radar range equation,*

$$r_{\max} = \frac{1}{2\sqrt{\pi}} \sqrt[4]{[P_t T] \times \left[\frac{1}{W_{r_{\min}}}\right] \times [G_t A_r] \times [A_o]} \tag{6}$$

Energy content of the transmitted pulse — Minimum detectable echo-pulse energy — Characteristics of antenna — Reflecting effectiveness of object

The bracketed factors in Eq. 6 show the effect of radar-set characteristics and the object on maximum range. The effectiveness of the trans-

mitter in making r_{max} large is measured by P_tT, the energy of each pulse transmitted. The effectiveness of the receiver depends on $W_{r_{min}}$, the minimum detectable echo-pulse energy. Antenna effectiveness is measured by the power gain G_t for the transmitting antenna and the effective area A_r for the receiving antenna. If, as is usual, the same antenna is used for transmitting and receiving, the subscripts t and r may be omitted, and the over-all antenna effectiveness then depends on the product GA. Large changes of any one of the bracketed factors produce only small changes of maximum range because of the fourth root in Eq. 6. If, for example, peak transmitter power is doubled, the range is increased by only 19 per cent (the fourth root of 2 is 1.187).

Minimum detectable energy $W_{r_{min}}$ depends largely upon the receiver, especially the critical input circuits. Relatively common inefficiencies of the input circuit may increase $W_{r_{min}}$ so much as to cause an important loss of range, despite the effect of the fourth root. Other parts of the radar system have minor effects upon $W_{r_{min}}$. Thus the indicator affects $W_{r_{min}}$ because the visibility of an echo pulse against the noise background is influenced by the type of data presentation employed. The pulse repetition frequency affects $W_{r_{min}}$ because it influences the appearance of noise and echo pulse on the indicator screen. Within limits, increasing the repetition frequency can make a weaker echo pulse visible. For this reason, the average power output of the transmitter (which is proportional to repetition frequency if the pulse energy is fixed) has an effect upon maximum range. The rate at which the antenna scans the region around the radar set affects $W_{r_{min}}$ because the number of echo pulses that are received from an object as the beam passes over it influences the visibility of the resulting signal. This number of pulses depends on the antenna beam angle and the repetition frequency, as well as on the rate of antenna scan.

If the same antenna is used for transmission and reception, the power gain G and the effective area A are related by the equation[1]

$$G = 4\pi \frac{A}{\lambda^2} \tag{7}$$

Therefore the factor GA is proportional to A^2/λ^2, and at fixed wavelength the maximum range is proportional to the *square* root of effective antenna area. For antennas of the same type but different size, effective and actual areas are proportional, and therefore maximum range is directly proportional to the linear dimensions of the antenna if the other bracketed factors in Eq. 6 are constant. Increasing the size of the antenna has,

[1] Equations resembling Eq. 7 but with constants different from 4π are often seen. These equations relate gain to actual antenna area instead of to effective area.

however, the additional effect of decreasing the beam angle. The increase of range resulting from an increase of antenna dimensions may therefore be offset by the smaller number of pulses returned from each object. In fact, if the antenna is made very large, the beam may become so narrow that objects are missed altogether—the antenna may rotate past the objects during the intervals between pulses.

6. Limitations of Radar. The performance of a radar set—its ability to measure range and direction angles of objects in its vicinity—is limited in several respects. Objects at great range return echoes too weak to produce discernible signals. Echoes from closely spaced objects merge and cause several objects to appear as a single one. The accuracy of measurement of range and angles is limited, and so is the speed with which a volume of space can be explored in the search for objects. Knowledge of the limitations of radar is necessary for the proper application of radar sets and for an understanding of radar components. Consequently a discussion of the more important limitations is worth while.

Maximum Range. For a radar set and object isolated in space, the maximum range of detection depends, according to Eqs. 6 and 7, on the energy of each transmitted pulse, on the minimum detectable echo-pulse energy, on the effective area of the antenna relative to the square of the wavelength, and on the radar cross section of the object. The minimum detectable echo-pulse energy depends mainly on the quality of the receiver input circuits and to a smaller extent on the over-all bandwidth of the receiver, on the type of data presentation employed, on the pulse repetition frequency, and on the number of pulses returned from the object as the antenna beam scans past it.

In actual operation of a radar set, the maximum range attained for a given object may be greatly different from the free-space maximum range because of the effects of the surface of the earth and the atmosphere. The actual maximum range is, however, always increased or decreased by changes in the radar set in the same way as is the free-space maximum range—except that changes in radar-set characteristics have very little effect when range is limited by the radar horizon.

Maximum range is limited in another way if the echo signal is expected to be observed at correct range on the indicator screen. In Fig. 13, echo pulse A_1 is returned by an object at short range with the delay time Δt_1 measured from the transmitted pulse A. Since Δt_1 is less than Δt_{max}, the time interval corresponding to the maximum range of the indicator sweep, echo pulse A_1 causes a signal at the correct range in the indicator display. An object at a range greater than the maximum range of the sweep, corresponding to echo pulse A_2, causes no signal on the indicator because the screen is blanked (see Art. 9) at the time of pulse A_2. An object at still greater range may return an echo pulse such as A_3 at a time after the

beginning of the next repetition period. Such a pulse is received during the sweep trace of the following repetition cycle and produces a signal in the indicator display, but at a range much less than the actual range of the object.

In the design of a radar set, the maximum range of the indicator sweep is chosen large enough to indicate objects at the greatest range expected to be useful, and the repetition period is made long enough so that echo pulses corresponding to A_3 in Fig. 13 are too weak under normal conditions to cause a discernible signal. The maximum range of a radar set

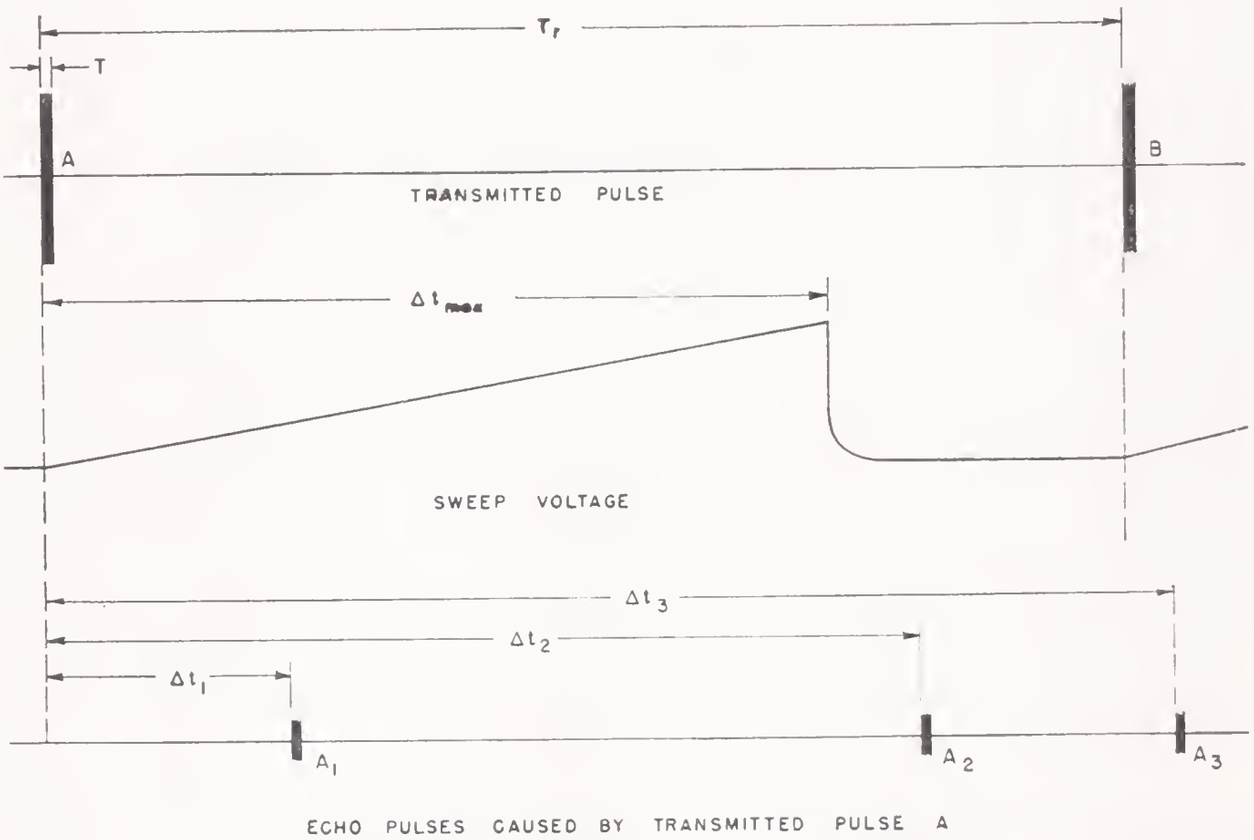


FIG. 13. Delay of echo pulses from objects at ranges greater than the maximum range of the indicator sweep.

therefore places an upper limit on the repetition frequency that can be used. If, for example, 200 miles is the greatest range from which detectable echoes are expected, the repetition frequency must not exceed $1/(200 \times 10.74 \times 10^{-6}) = 497$ cps. The maximum time duration of the indicator sweep must be less than the repetition period, and therefore the maximum sweep range must be less than 200 miles if 497 cps is chosen as the repetition frequency.

Under certain atmospheric conditions, echoes are returned from large objects at ranges far beyond the normal maximum. Many radar sets therefore occasionally receive echo pulses delayed by more than one repetition period, and radar operators must be aware of this possibility.

Minimum Range. Objects at very short range, as well as those at extremely great range, cannot be detected by radar. One reason for the

minimum-range limitation is that most transmitters cannot end a pulse with perfect suddenness. For several microseconds following the main part of the pulse, relatively weak signals may be produced (Art. 11, Chap. XI). These signals are too weak to cause useful echoes but are strong enough to mask echoes from near-by objects. Another effect that limits minimum range is the recovery time of the t-r device. During the transmitted pulse, the t-r device must block signals from the receiver input with great effectiveness. Following the pulse, a short time interval is required for the gas in the t-r tube to deionize and allow signals to reach the receiver. As deionization progresses, the sensitivity of the receiving system is restored gradually during the first few microseconds following the transmitted pulse. Another possible limit on minimum range is receiver recovery time. Despite the t-r device, the transmitted pulse causes a receiver-input signal much larger than normal echo signals, and this large signal drives amplifier circuits in the receiver beyond their normal range. The minimum range of detection is increased if an appreciable time is required for the circuits to recover normal sensitivities after this overload.

Receiver recovery time can be made negligible by proper design, but the limitation of the t-r device recovery time and the transmitter output following the pulse remain. The actual value of the minimum range depends, like the maximum range, not only on the radar set but also on the radar cross section of the object. Larger objects can be detected at shorter ranges because the greater echoes from them can be discerned despite the interfering signal from the transmitter and the attenuation in the t-r device.

Minimum Cross Section of Object. Because echo-pulse energy depends not only on range but also on reflecting properties of the object, an object at a fixed range can be detected only if its radar cross section exceeds some minimum value. The minimum cross section increases as range increases, in accord with the radar range equation, and also increases at very short ranges because of the minimum-range limitations. The variation with range of the minimum cross section for detection is represented by the curve of Fig. 14. The minimum and maximum ranges at which an object of given cross section can be detected can be found on this curve as

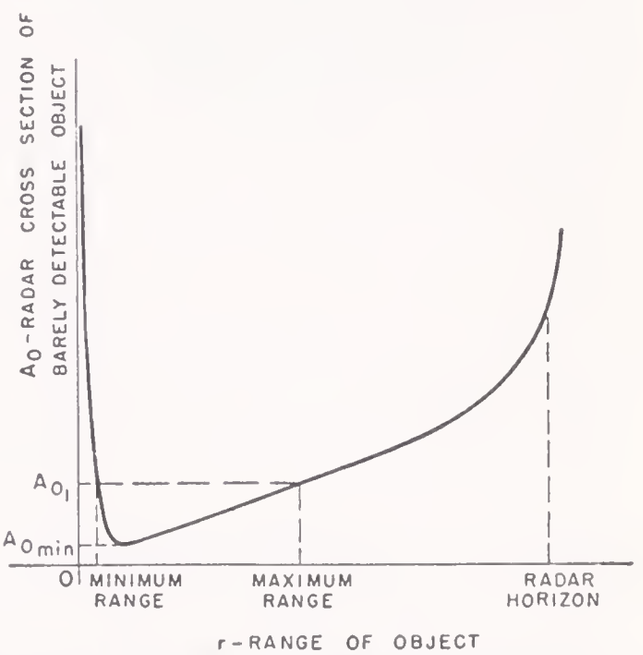


Fig. 14. Relation of radar cross section to range for a barely detectable object.

illustrated for an object of cross section A_{o_1}. Note that an object with cross section less than $A_{o_{min}}$ cannot be detected *at any range.*

From the radar range equation (Eq. 6) it may be seen that doubling the energy of the transmitted pulse has the same effect on range of detection as does multiplying the ordinates of the curve in Fig. 14 by 2, thus increasing very slightly the range at which large objects can be detected and increasing considerably the detection range for small objects. Similarly, halving the minimum discernible echo-pulse energy halves the ordinates of the curve, and doubling the antenna gain and effective area (see Eq. 6) divides the ordinates of the curve by 4.

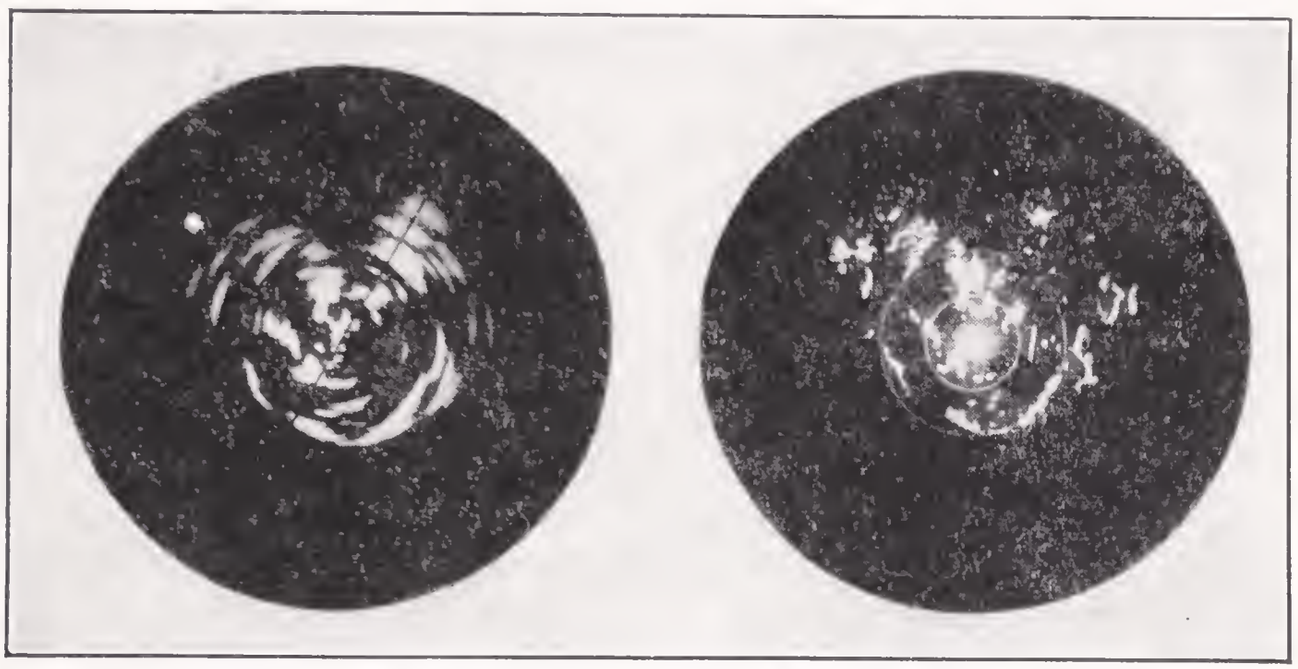
The over-all ability of a radar set to detect objects depends upon (1) the ratio of transmitted-pulse energy to detectable echo-pulse energy and (2) the antenna gain. Antenna gain is ordinarily constant for a given radar set, but the ratio of pulse energies may vary widely, because it depends on the condition of the transmitting and receiving circuits. The value of the pulse-energy ratio therefore indicates whether or not the ability of the radar set to detect objects is the maximum possible. This ratio is sometimes called the *over-all performance* of the radar set. For a set that transmits pulses having a peak power of 100 kw and is capable of detecting pulses of the same duration as the transmitted pulse and a peak power of 1 $\mu\mu$w, the over-all performance ratio is 100,000,000,000,000,000.

Because maximum range on large objects is usually horizon-limited and is therefore little affected by radar-set performance, radar sets may seem to be operating satisfactorily even though the over-all performance ratio is far below normal and range on small objects is therefore reduced. Test equipment capable of measuring over-all performance is essential for proper maintenance of radar sets. An especially simple and useful testing device is the *echo box* (Art. 18, Chap. XII) with which relative over-all performance may be determined—the actual pulse-energy ratio is not usually measured, but an indication of the actual over-all performance relative to its normal value is obtained.

Resolution. The maximum-range, minimum-range, and minimum-cross-section limitations relate to the possibility of an object returning an echo pulse strong enough to be discernible. Obtaining useful information from echo signals requires, in addition, reception of separate and distinct signals from the various objects surrounding the radar set. When signals from several objects merge, individual objects often cannot be identified, and the positions of objects cannot be determined with accuracy. The relation of resolution to the amount of information in a radar display is illustrated by the photographs of two plan-position presentations in Fig. 15.

Range resolution depends principally on the pulse duration of the transmitter (see Art. 1), but it may be reduced by the receiver bandwidth,

because a narrow-band receiver tends to spread out the pulses it amplifies. Ordinarily, radar receivers have bandwidths great enough so that the spreading effect is negligible, and the transmitter pulse duration alone determines the range resolution. Angular resolution depends on the beam angle of the antenna (see Art. 2) and may be different in azimuth and elevation if the antenna beam is narrower in one direction than the other. Because the edges of an antenna beam are not sharp, the setting of the receiver gain control also influences angular resolution. The arc length corresponding to a given beam angle is proportional to range. At long ranges, therefore, the resolution of a display is limited mainly by the beam angle and at short ranges by the pulse duration. Suppose, for



(*A*) Low-resolution presentation of a 10-cm radar set. (*B*) High-resolution presentation of the same area on a 1-cm set.

FIG. 15. Photographs of plan-position presentations illustrating the effect of resolution.

example, that the pulse duration is 1 μs and the azimuth beam angle is 1 deg. The spread of object indications in range is then 492 ft. The spread in azimuth is also 492 ft *if the range is* 5.32 *miles* and is more at longer ranges, less at shorter ranges.

An additional limitation on resolution that might be considered is the diameter of the spot on the cathode-ray tube. A large spot would cause indications from adjacent objects to merge. This limitation is not fundamental, however, because expanded displays can be made that spread small sections of the range and azimuth scales over the entire indicator screen. Such displays are used whenever the resolution otherwise obtainable warrants them.

The obscuring of echoes by ground clutter and sea return is closely related to resolution. If the antenna beam of a ground radar set is wide in elevation, echo signals from an airplane may be lost among echoes at

the same range and azimuth from the surrounding terrain. Air-borne radar sets cannot ordinarily detect ships directly below themselves because of the tremendous echo returned from the surface of the sea. Echoes from ship and sea cannot be separated because the ranges differ by less than the pulse duration. In general, the disturbing effects of any type of interfering signal are reduced by an increase in resolution, because the region of the display affected by the interference is decreased in size.

Accuracy. Detection of objects by radar is not enough in many applications; range, azimuth, and elevation coordinates of the objects must be measured with considerable accuracy. Because the requirements for accurate range measurement are different from those for accurate angle measurement, the two are discussed separately.

Comparison of echo signals with a range scale on a Type A, B, or P indicator in the manner described in Arts. 1 and 3 yields only approximate values of range. Many applications require greater accuracy, and for these a number of high-accuracy range circuits have been developed. To show which characteristics of a radar set limit the attainable accuracy, it is necessary first to describe the principle of one method of accurate range measurement.

Assume that the range of the strongest echo signal within the dotted circle drawn on the Type A presentation of Fig. 16*A* is to be measured. An expanded presentation of the portion of the trace within the circle is made as in Fig. 16*B*. To obtain the expanded display, the start of the sweep is delayed by a controllable amount and the sweep is made very fast. Note that the short echo pulses that appear as "spikes" in Fig. 16*A* are spread out in Fig. 16*B* so that the shape of the pulses is evident. A voltage is added to the signal voltage on the vertical-deflection plates to produce a *range step* (or sometimes a notch or other marker waveform), as shown in Fig. 16*C* at a fixed time interval after the start of the expanded sweep. Adjusting the delay of the sweep and step causes the train of echo pulses to move relative to the step. The delay is adjusted until the *leading edge* of the desired echo pulse coincides with the step, and the range is then read from the dial that controls the delay.

The accuracy of the range measurement depends on the circuits that control the delay of the range step, on the setting of the edge of the echo pulse to coincide with the step, and on the relation of range to the propagation time of radio waves. Range circuits can be made with almost any desired accuracy but become very complex when exceedingly small errors are required. The accuracy of setting the echo pulse is limited by the steepness of the leading edge of the pulse.

The edge of the echo pulse cannot be steeper than the edge of the transmitted pulse, and therefore, a very sharp rise of the transmitted pulse is essential to range accuracy. The main reason for using the leading

edge in range measurements is that the leading edge of the transmitted pulse can be made steep much more easily than can the trailing edge (Arts. 8 and 10, Chap. III). Even if the transmitted-pulse edge were infinitely steep, the echo-pulse edge would slope because of the limited bandwidth of the receiver; the rate of rise of the echo-pulse edge would be proportional to receiver bandwidth. Consequently, accurate range

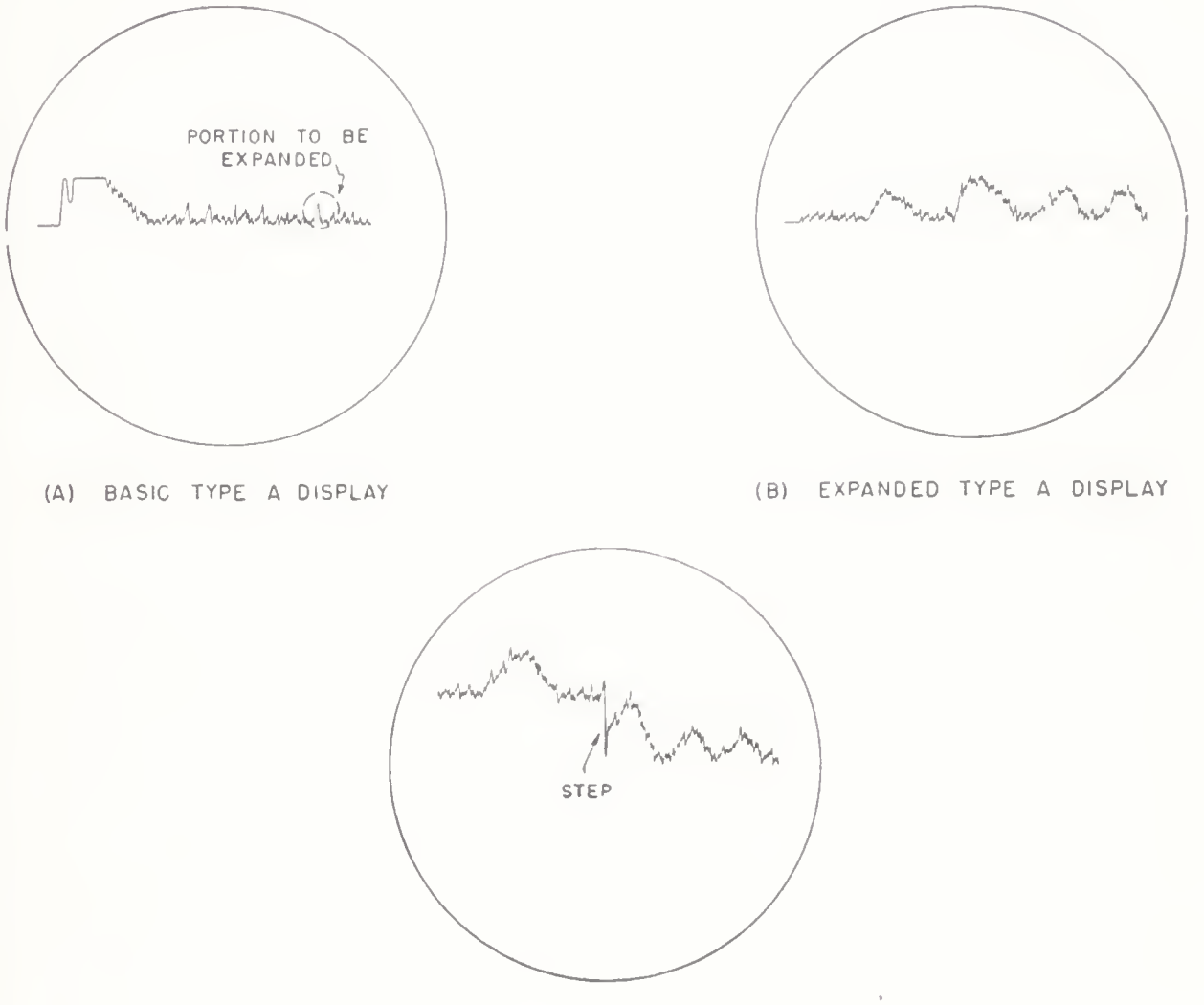


FIG. 16. Indicator presentation used in one method of measuring range accurately.

measurement requires the greatest possible bandwidth—a requirement in conflict with the bandwidth for best discernibility of the echo signal.

The transmitted-pulse shape and the receiver bandwidth determine the time required for the echo-pulse leading edge to rise to a fixed percentage of its maximum amplitude. The actual rate of rise of the pulse edge depends also upon the strength of the echo. Therefore, if the echo pulse is weak so that the usable receiver gain is limited by noise, range accuracy is dependent on echo-pulse amplitude. If, however, a strong echo is available so that the noise amplitude is negligible, echo-amplitude changes can be compensated for by adjustment of the gain control and have no effect on range accuracy.

To determine angular coordinates of an object, the position of the antenna when maximum echo signals are received may be observed, as explained in Arts. 2 and 3. Angular errors are then of the same order of magnitude as the beam angle of the antenna. For a specified beam angle, greater accuracy can be had if *lobe switching* is employed. The antenna is constructed so that its beam can be made to occupy either of two positions separated by an angle of about the same size as the beam angle, and the beam is switched from one position to the other several times a second. Echo pulses returned when the beam is in each of the two positions produce separate signals on the indicator, and the antenna is positioned so that the two signals are of equal amplitude. The direction of the reflecting object is then midway between the two positions of the axis of the antenna beam. (For further discussion of lobe-switching methods, see Art. 9, Chap. IV.) Lobe switching increases the accuracy of angle measurement because intensity of the radiation from the antenna changes more rapidly with angle near the edge of the beam than at the maximum point.

A motion of the antenna beam called *conical scan*, which is often used for the accurate determination of angle, may be regarded as a form of lobe switching. An antenna beam of circular cross section is moved continuously in such a way that its axis describes a cone having an angle at its apex about equal to the beam angle. If an object lies on the axis of the cone, the rotary motion of the antenna beam does not cause the strength of the returned echo to change, because the angle between beam axis and a line to the object is constant. If the object lies to one side of the cone axis, the echo strength increases and decreases during each tracing of the cone. To measure the angular position of the object, the antenna is turned until the conical scan causes no variation of echo strength. The object direction is then defined accurately by the axis of the cone. Conical scan is equivalent to simultaneous lobe switching in azimuth and elevation.

The use of the edge of the antenna beam in an angle measurement may be compared with the use of the edge of a pulse in the measurement of range. Because the accuracy of either measurement is limited by the sharpness of the edge, range errors much less than the range equivalent of pulse duration and angle errors less than beam angle are obtainable. Observe, however, that such accuracy can be had only with relatively isolated objects. When several objects lie close together, high resolution is necessary to separate the signals they produce before the coordinates of any one of them can be determined with accuracy.

Speed of Collecting Data. When objects move relative to a radar set, the set is required not only to detect the objects, to indicate the objects separately, and to determine the coordinates of each but also to perform

these functions again and again with great rapidity. The speed of the process is especially important when the courses of fast-flying aircraft are to be followed; the successive instants at which the positions of the aircraft are indicated must be closely spaced to limit the distances traveled in the intervening intervals.

The beam from a radar antenna is required to scan a definite region of space, that is, to explore all parts of the region in search of reflecting objects. The position of any one object is indicated when the beam is directed toward the object, and this occurs once in each complete scanning cycle. Thus a rapid scan is required if adequate data on fast-moving objects are to be obtained. The rate of scan is limited, however, because several pulses must be transmitted while the beam is passing across any one object. Otherwise, discernible signals would not be produced by the echo pulses. The allowable rate of antenna scan depends, therefore, on repetition frequency and beam width. Increasing either of these quantities permits an increase in the rate of antenna motion without a loss in the number of echo pulses returned by each object.

Because broad antenna beams cannot be used when high angular resolution is required, speed of data collection must often be sacrificed in high-resolution sets. In certain applications, the speed can be increased by proper shaping of the antenna beam. For example, if a pencil beam (a narrow beam of circular cross section) is used to search for aircraft, the sky may be scanned by rotating the antenna rapidly about a vertical axis and tilting it slowly in elevation, but the complete scanning process requires a long interval of time. In air-search sets, the antenna beam is usually fan-shaped—narrow in azimuth but broad enough in elevation[1] so that the entire region is scanned by a single revolution of the antenna. The time required for the scanning process is reduced, and yet good accuracy and resolution in azimuth are retained.

For a given beam angle, the allowable rate of scan depends on repetition frequency. Therefore (since the maximum range of the set depends on the energy of each transmitted pulse), the product of range and scan rate, which is a measure of the data-collecting speed, depends on the product of repetition frequency and pulse energy, or the average power of the transmitter. Data can be collected most rapidly, therefore, if large, heavy, expensive transmitting equipment is employed to provide high average power. An ultimate limit is reached, however, when the repetition frequency is increased to the maximum value that permits two-way travel time for a maximum-range echo. The speed of radio waves may, therefore, be considered to place an ultimate limit on the speed at which a radar set can collect data.

The requirements placed on characteristics such as pulse duration,

[1] The cosecant-squared pattern is often used. See Art. 15, Chap. XIII.

bandwidth, and beam angle by desired performance features such as range, resolution, and speed of data collection are often conflicting. For example, if small objects are to be detected at long range, a long pulse should be used to provide high pulse energy, the repetition frequency should be low to minimize the average power and prevent range ambiguity, the antenna should have high gain and should rotate so slowly that each object is illuminated by a long succession of pulses, and the receiver bandwidth should be correctly related to the pulse duration. Use of a long pulse is, however, in conflict with the requirement for range resolution; a higher repetition frequency, broader antenna beam, and more rapid antenna-scanning motion should be used to obtain data more rapidly; and a greater receiver bandwidth should be employed for accurate determination of range. These conflicting requirements are the reason that radar sets are made with a variety of different characteristics, each suited to its own application.

7. Radar Frequencies. The frequencies of the radio waves employed in radar lie in the general region of 100 to 30,000 Mcps. How extremely high these frequencies of billions of cycles per second are may possibly be suggested by the comparison of power, communication, radar, and

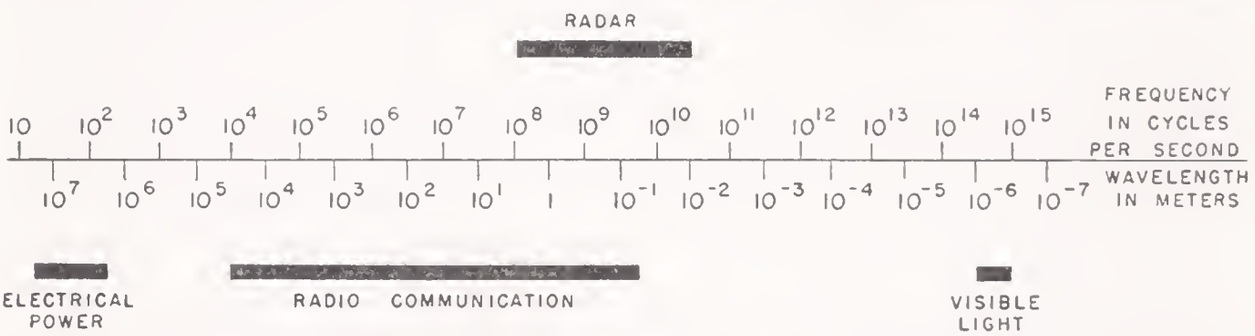

RADAR
10 10² 10³ 10⁴ 10⁵ 10⁶ 10⁷ 10⁸ 10⁹ 10¹⁰ 10¹¹ 10¹² 10¹³ 10¹⁴ 10¹⁵
FREQUENCY IN CYCLES PER SECOND
10⁷ 10⁶ 10⁵ 10⁴ 10³ 10² 10¹ 1 10⁻¹ 10⁻² 10⁻³ 10⁻⁴ 10⁻⁵ 10⁻⁶ 10⁻⁷
WAVELENGTH IN METERS
ELECTRICAL POWER
RADIO COMMUNICATION
VISIBLE LIGHT

FIG. 17. Comparison of radar frequencies with power, communication, and light frequencies.

light frequencies in Fig. 17. On the logarithmic scale of this figure, radar frequencies lie closer to visible light than to electric-power frequencies. It is not strange, then, that much radar equipment (for example, the antenna with paraboloidal reflector in Fig. 6) resembles optical devices more closely than low-frequency electrical equipment.

Radar wavelengths range from 1 cm to 3 m in contrast to 200 to 550 m for the low-frequency radio-broadcast band and 0.00004 to 0.00008 cm for visible light. The relation between frequency and wavelength, indicated by the scales in Fig. 17, is determined as follows: A wave of frequency f moves forward one wavelength λ in each cycle of oscillation. Therefore, the speed with which the wave progresses, which is the distance moved forward in one unit of time, is $f\lambda$. For radio waves, this speed is c, the speed of light, 299.8 m per μs. Therefore,

$$f\lambda = c \tag{8}$$

$$(\text{Mcps}) \times (\text{m}) = (\text{m per } \mu\text{s})$$

The units indicated below Eq. 8 are convenient for use at radar frequencies. If the correspondence of 1 m to 300 Mcps (more accurately 299.8 Mcps) is remembered, the inverse proportionality between frequency and wavelength can be used for any desired frequency–wavelength conversion.

The range of frequencies used by radar sets are sometimes divided into bands designated by the letters *P*, *L*, *S*, *X*, and *K*. These bands are centered on wavelengths of approximately 2 m, 30 cm, 10 cm, 3 cm, and 1 cm, respectively.

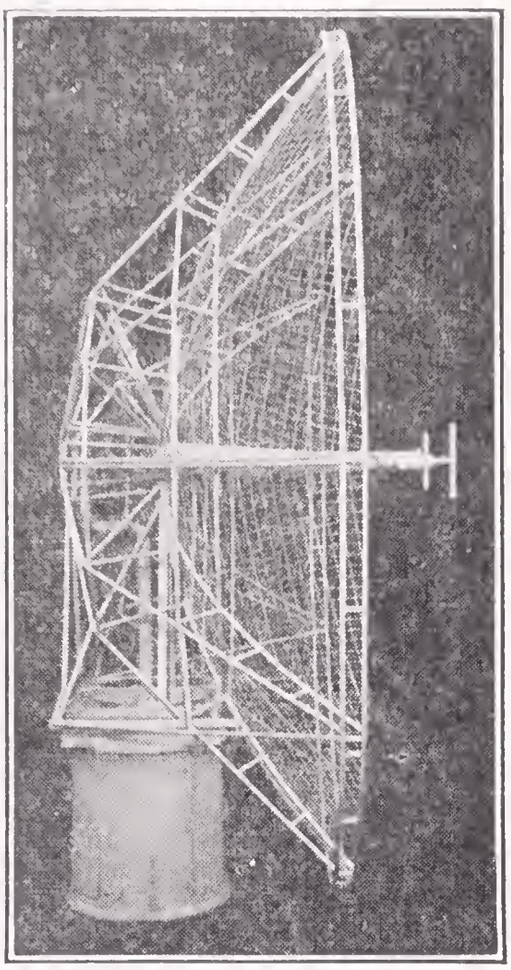


FIG. 18. Example of a *P*-band (operating frequency approximately 220 Mcps) paraboloidal antenna. The antenna diameter is 17 ft; the beam angle is 20 deg. (*Courtesy of General Electric Company.*)

Radar equipment operating in the *P* and *L* bands is known as *ultrahigh-frequency* (abbreviated uhf) radar, and that in the *S*, *X*, and *K* bands is called *microwave* radar. The distinction is of importance because several characteristic differences exist between uhf and microwave equipment. Principally, the antenna, transmission lines, transmitting tubes, and receiver input circuits take different forms for different frequencies, and the performance characteristics of the set depend greatly on the frequency of operation.

Typical uhf and microwave antennas are illustrated in Figs. 5 and 6. The paraboloidal reflector and feed of Fig. 5 is a microwave antenna, and the array of half-wave elements mounted on a reflecting screen shown in Fig. 6 is used at ultrahigh frequencies. (The various antennas mentioned here are discussed in detail in Chap. XIII.) Paraboloidal reflectors are used in some uhf sets, especially at frequencies near 1,000 Mcps. Ultrahigh-frequency paraboloids are very large and are made of wire, rods, or other open metalwork, as illustrated in Fig. 18, in order to reduce weight and wind resistance. Another form of uhf antenna is the Yagi array (see Art. 12, Chap. XIII). Many special antenna types are found in microwave sets. Examples are the truncated paraboloid of Fig. 7, the parabolic cylinder of Fig. 19, and the end-fire array of Fig. 20. Because of the advantage of short wavelength in producing directivity, microwave antennas tend to be smaller than uhf antennas and to produce narrower beams. Beam angles of 2 to 30 deg may be expected with uhf antennas, while 0.1 to 10 deg is typical of microwave antennas.

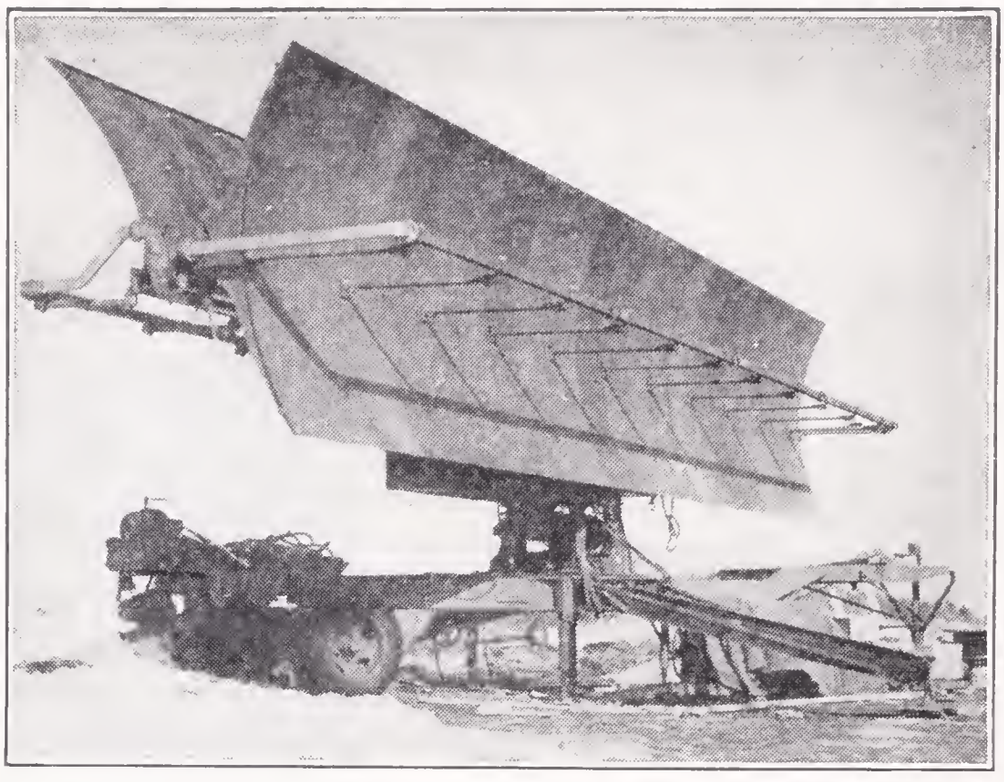


FIG. 19. Example of a parabolic-cylinder antenna.

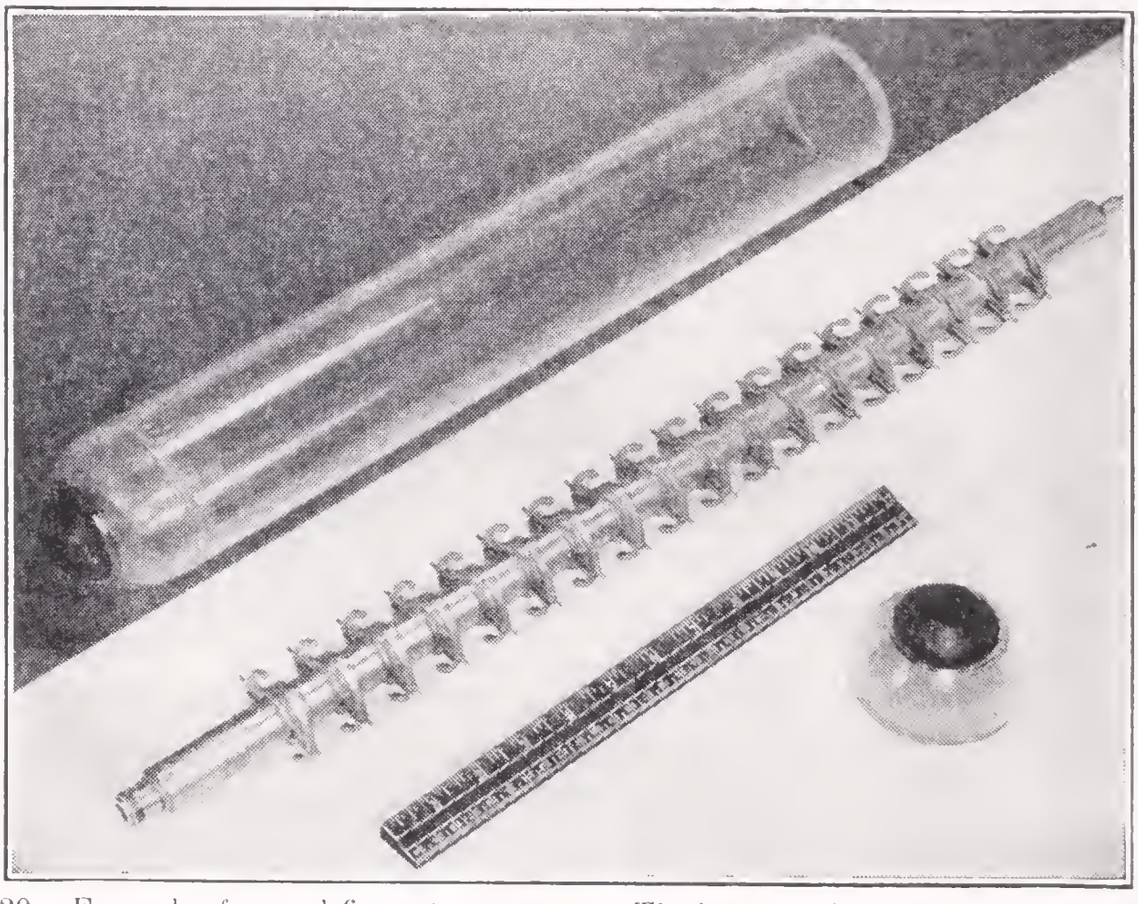
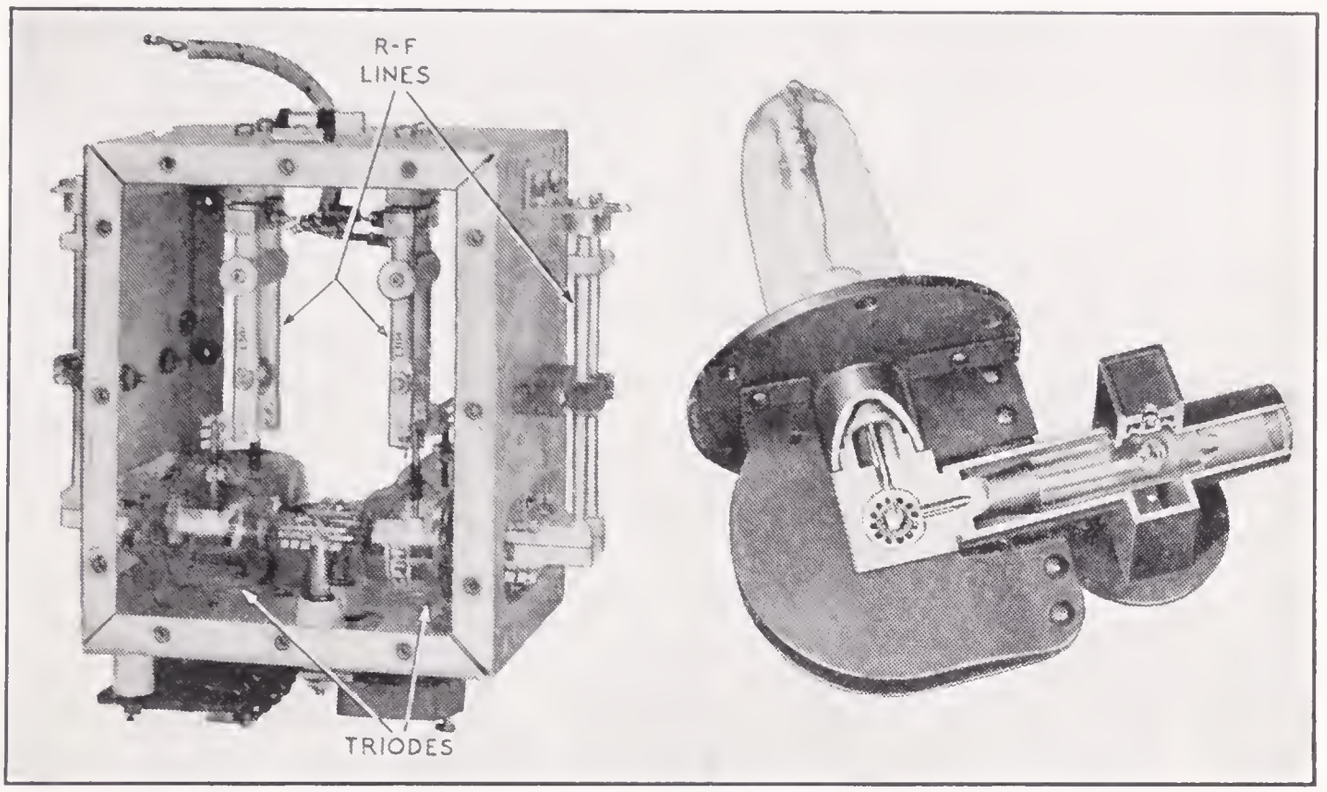


FIG. 20. Example of an end-fire antenna array. The beam angle is approximately 28 deg.

Transmission lines (see Chap. VII) in uhf radar equipment are often of the *parallel-wire* type, similar to the transmission lines used for power and communications at low frequencies. *Coaxial* lines also are frequently employed, especially at the high-frequency end of the uhf region. In these, one conductor is a rod placed along the axis of an outer cylindrical

conductor. These transmission lines conduct r-f power from transmitter to antenna and from antenna to receiver, and short sections of line are used as tuning elements throughout the r-f portions of the radar set. Microwave equipment typically employs *waveguides* (Chap. VIII) instead of transmission lines, although coaxial lines are used in low-power and medium-power sets at 10 cm. Tuning elements in microwave sets are often *resonant cavities* (Chap. IX), many of which may be regarded as sections of waveguides or coaxial lines.

Transmitting tubes of two types are used in radar. One type is the *triode* (Chap. X), especially designed for high-power, high-frequency



(*A*) 500-Mcps transmitter. (*Courtesy of Radio Corporation of America.*)

(*B*) Sectional view of Type 725A (3-cm) magnetron. (*Courtesy of Bell Telephone Laboratories.*)

FIG. 21. Comparison of uhf and microwave transmitting tubes and tuning elements.

operation, and the other type is the magnetron (Chap. XI). Triodes can produce large amounts of power in the uhf region but have much lower output than the magnetron if they must operate in the microwave region. Magnetrons, on the other hand, are practical in the microwave bands but are very large and heavy if designed for lower frequencies. Consequently, uhf sets employ triode transmitting tubes, and microwave sets use magnetrons. Exceptions exist, however: 700-Mcps magnetrons have been used, and *disk-seal triodes* (Art. 1, Chap. X) are suitable as transmitting tubes in low-power 10-cm sets. Figure 21*A* shows the internal construction of a typical uhf transmitter. The triode tubes are in evidence, as are the transmission-line sections used as tuning elements. Figure 21*B* shows, for comparison, the construction of a 3-cm magnetron

transmitting tube. The magnetron makes use of resonant cavities as tuning elements.

Microwave and uhf receivers differ principally in the input stages. Because of the difficulty of obtaining amplification at high frequency, substantially all radar receivers employ a superheterodyne circuit in which the echo pulses received are changed to pulses of lower frequency and then amplified. Stages operating at the lower frequency can be similar in uhf and microwave receivers. The change of frequency

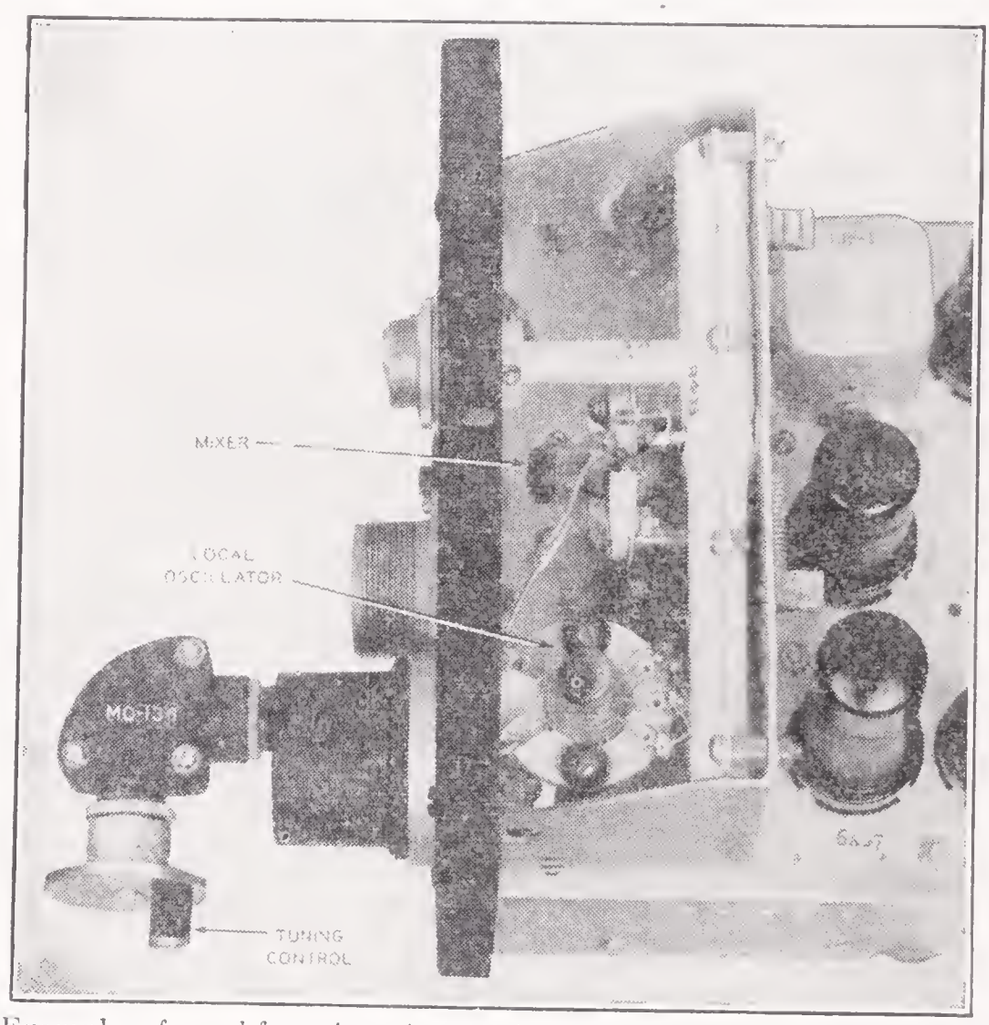


FIG. 22. Example of a uhf-receiver input circuit. (*Courtesy of Radio Corporation of America.*)

requires the use of a *local oscillator* to generate a signal of frequency nearly equal to that of the received pulse and a nonlinear mixer element to rectify the combination of local-oscillator signal and echo pulse and produce a beat-frequency or difference-frequency pulse. The difference frequency is usually in the range of 15 to 60 Mcps and is called the *intermediate frequency* of the radar receiver. Microwave and uhf receivers employ different types of mixers and local oscillators, and in addition, uhf receivers commonly have one or two r-f amplifier stages before the mixer; in microwave receivers the input pulse is applied directly to the mixer. Special triodes, such as acorn tubes and especially lighthouse tubes (small disk-seal triodes), are useful in uhf sets as r-f amplifiers and local oscil-

lators and frequently as mixers. In microwave receivers, *crystals* (Art. 12, Chap. VI) are used as mixers and *klystrons* (Chap. XI) are commonly employed as local oscillators. Lighthouse-tube local oscillators are also used at 10 cm. The appearances of the input circuits of typical uhf and microwave receivers are shown by the photographs of Figs. 22 and 23. The sectional drawing of the klystron (Fig. 24) shows the resonant cavity that controls the frequency of the local-oscillator signal in the microwave

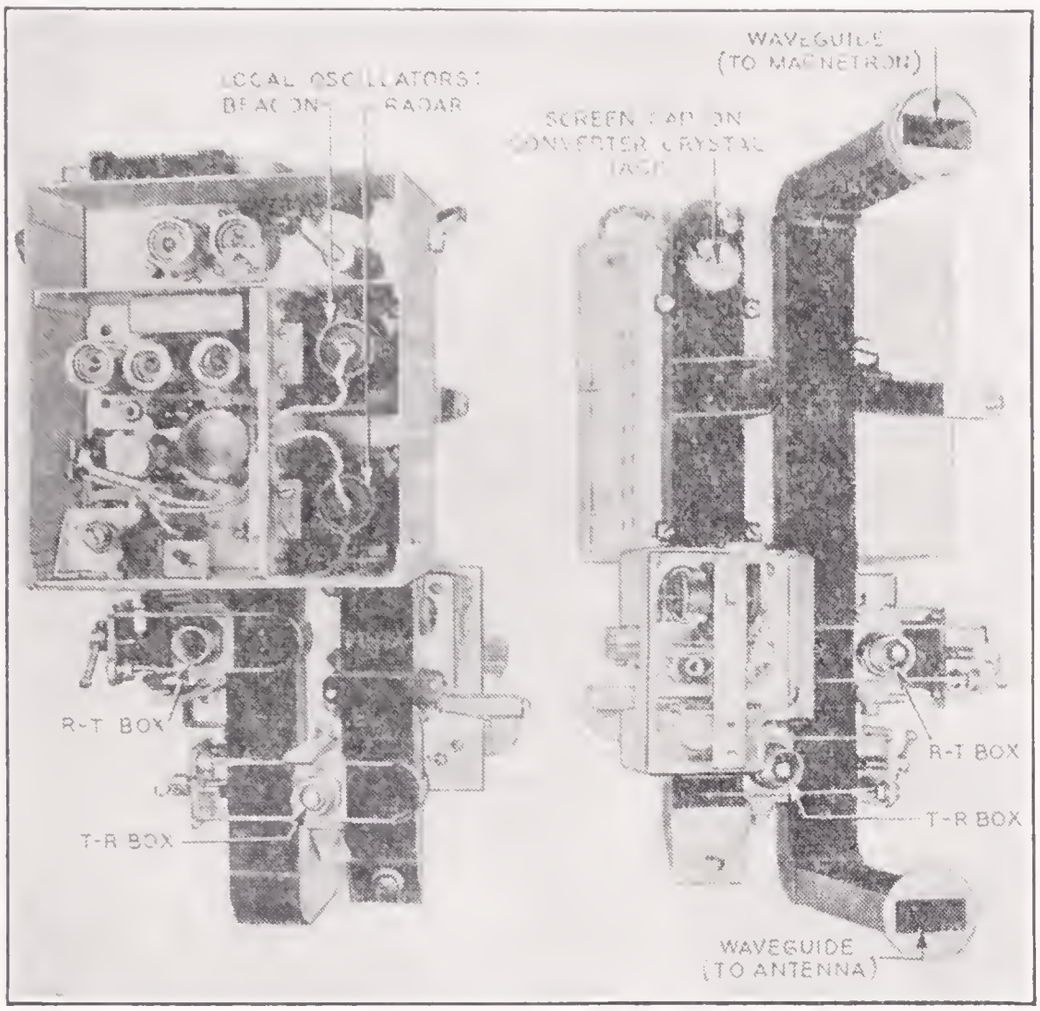


FIG. 23. Example of a 3-cm receiver input circuit. (*Courtesy of Bell Telephone Laboratories.*)

receiver. The features of typical microwave mixers may be seen in Figs. 54 and 55 of Chap. XII.

There are a number of reasons why radar frequencies are limited to the range of about 100 to about 30,000 Mcps. The lower limit is fixed partly by the difficulty of obtaining adequate antenna directivity at frequencies below 100 Mcps. The wavelength corresponding to 100 Mcps is 3 m, and at a lower frequency and correspondingly longer wavelength, an antenna having dimensions many times the wavelength would be impractical. Another effect that places a lower limit on radar frequencies is that little reflection is obtained from objects such as ships and aircraft if the radar wavelength is long relative to the dimensions of the object.

The upper limit for radar frequencies is fixed partly by the technical difficulty of generating and receiving waves of the highest frequencies. Transmitting tubes tend to produce less power, and the noise level in receiver circuits tends to increase as the frequency is raised. Tolerances

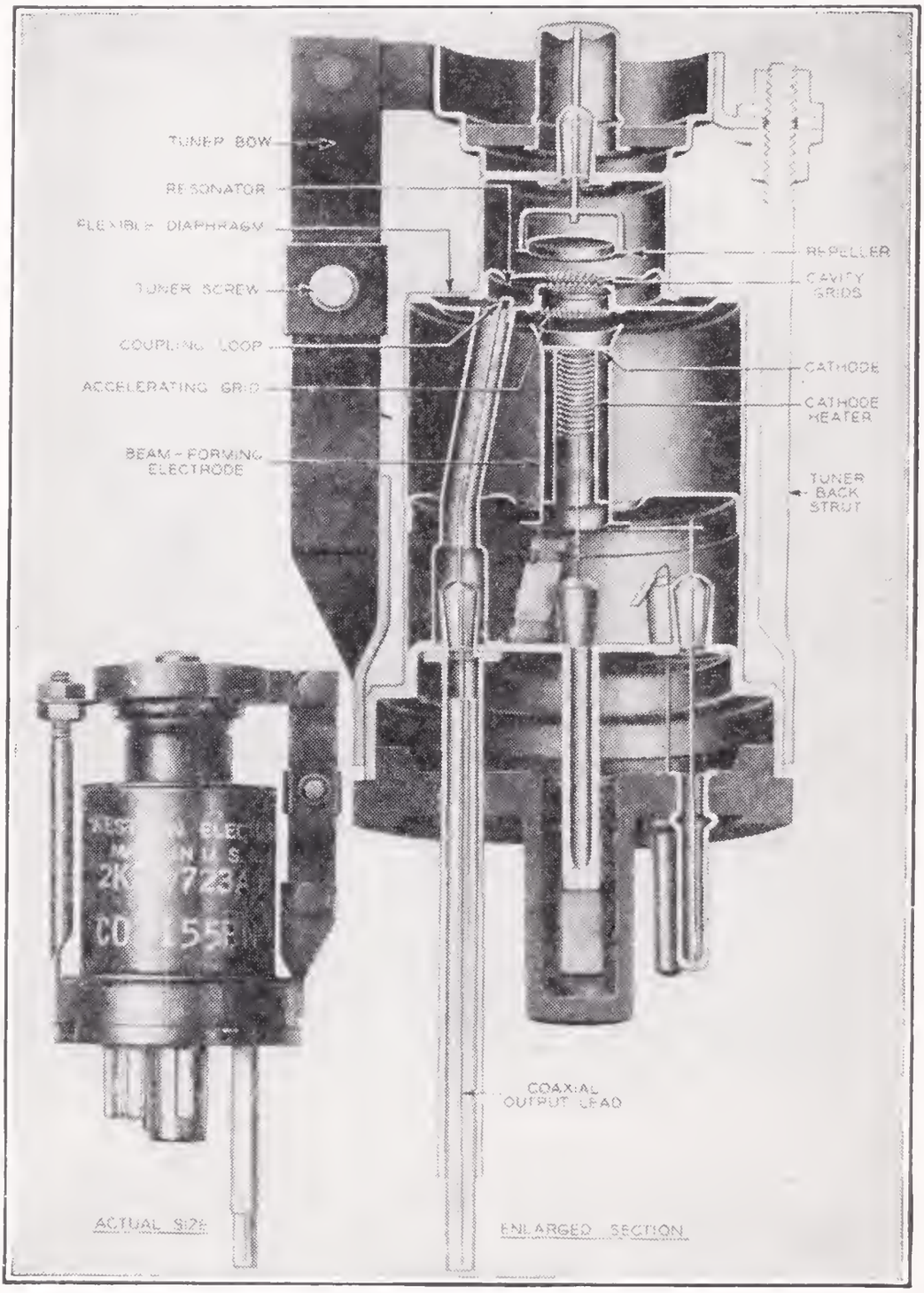


FIG. 24. Sectional view of klystron used in the receiver input circuit of Fig. 23. (*Courtesy of Bell Telephone Laboratories.*)

for dimensions of high-frequency components are a relatively fixed fraction of wavelength and thus become increasingly difficult to maintain as the frequency is raised.

The upper-frequency limit set by technical difficulties tends to rise as new tubes, circuits, and techniques are developed, but another effect seriously limits the usefulness of radar at frequencies of 30,000 Mcps or

higher: The absorption of radio waves by the atmosphere becomes important. Furthermore, fog, rain, and snow have little effect on low-frequency radio waves but cause considerable reflection and some absorption at high frequencies. Thus, one advantage of radar over optical detection methods is lost if too high a frequency is used.

The particular frequency within the range useful for radar that is best depends upon the use to which the radar set is to be put. In general, high radar frequencies are most suitable when high resolution or high accuracy is required. Somewhat lower frequencies are better if the longest possible detection range is the object, especially if the radar set is to be simple and inexpensive. The manner in which frequency affects radar performance may be summarized as follows:

Transmitter Performance. As the frequency is increased, the pulse energy that can be generated decreases and, therefore, the range attainable tends to decrease. An exception to this rule is that magnetrons can produce as large pulse energies at about 10 cm as can triodes at longer radar wavelengths; therefore, in terms of transmitter output, there is little advantage to frequencies below about 3,000 Mcps.

Receiver Performance. As the frequency is increased, there is a tendency for receiver noise to increase. For this reason also, attainable range tends to decrease with increase of frequency. Development of new circuits and techniques for reducing noise is one of the most promising possibilities for improving high-frequency radar.

Antenna Performance. The effect of frequency change on antenna performance depends upon whether or not the antenna size is changed simultaneously. If antenna size is not changed, increase of frequency narrows the antenna beam and increases the gain. The resolution of the set is therefore improved,[1] and the greater gain opposes the tendencies toward decrease of range with increase of frequencies. If the antenna size is reduced with increase of frequency in such a way that dimensions in wavelength remain constant, then directivity and gain do not change. The advantage of the higher frequency is then the decreased size of the antenna. The disadvantage is the reduced range, occasioned by reduced effective area of the antenna (see Eq. 7) as well as by the effects in transmitter and receiver.

8. Types of Radar Equipment. To meet the requirements of a host of different applications, radar sets having a wide variety of characteristics have been manufactured. Some are designed, for example, to detect objects at the greatest possible range. In others, range is sacrificed in order that accuracy and resolution may be made as high as pos-

[1] Antenna directivity controls only the angle resolution, but there is an associated increase in range resolution because somewhat shorter pulses may be used when the frequency is increased.

sible. Both range and accuracy are sacrificed in still other types, and the equipment is made especially simple, small, and lightweight. Many sets have auxiliary devices attached to them or are required to work in conjunction with nonradar equipment in order that tasks different from the basic radar functions may be performed. Some idea of the variety of radar sets may perhaps be had from a review of several methods of classifying radar equipment.

One classification scheme designates radar sets as *land stations, shipboard sets,* or *aircraft radar.* Fixed land stations are characteristically

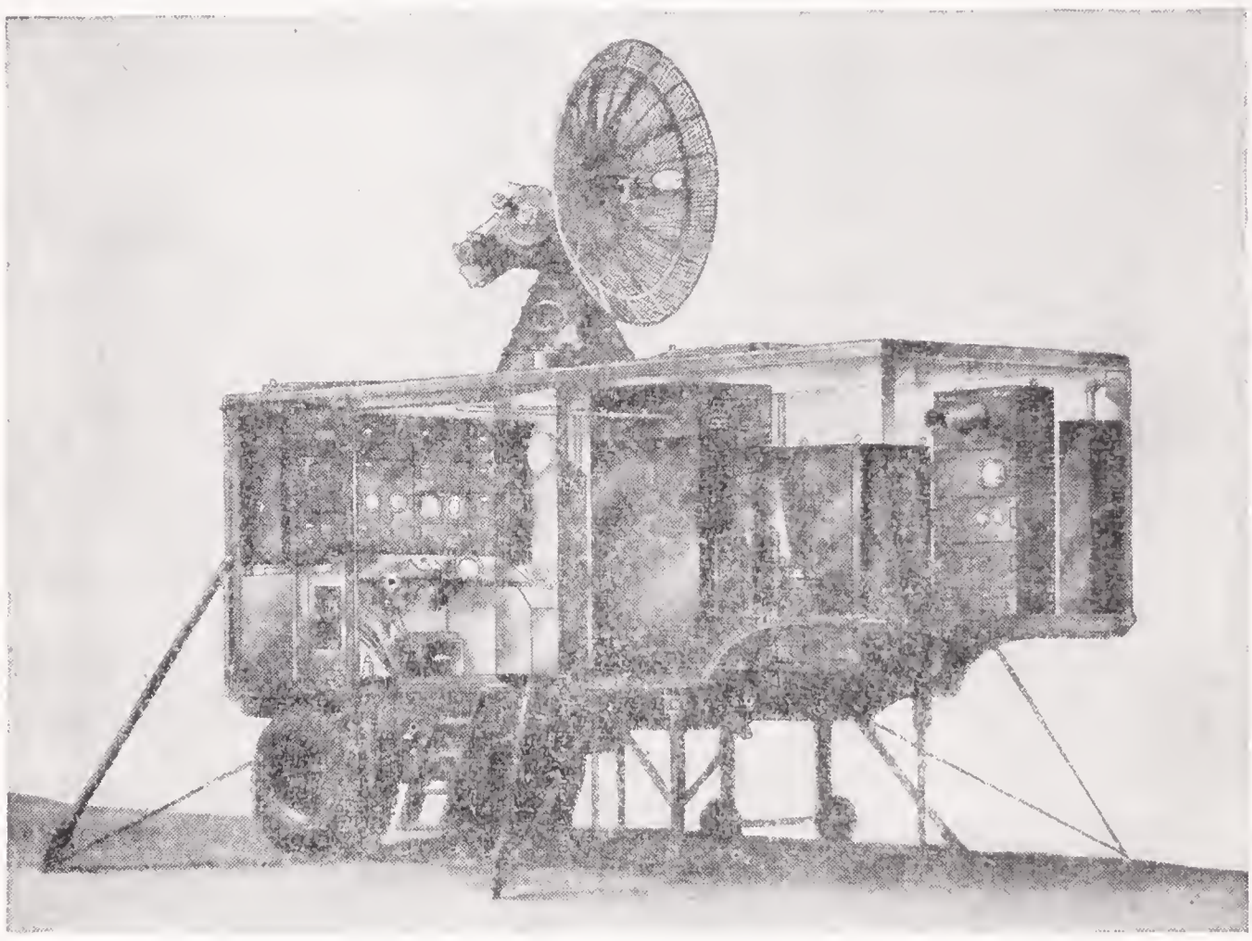


FIG. 25. Example of a mobile, land-based radar set (the SCR-584) used for air search and fire control. The operating frequency is approximately 3,000 Mcps, the pulse duration is 0.8 μs, the repetition frequency is 1,707 cps, and the peak power is 300 kw. [*Reproduced, with permission, from Louis N. Ridenour, Radar System Engineering* (*McGraw-Hill Book Company, Inc., New York,* 1947).]

large and complex and have tremendous antennas and powerful low- or medium-frequency transmitters. Very great ranges are therefore possible even with aircraft echoes, and resolution is moderate to excellent. Mobile land sets range from those which can be carried in several trucks to sets so small that one man can carry them. Reduced range and resolution are, of course, obtained from the smaller sets. An example of a mobile land-based set, the SCR-584, appears in Fig. 25.

Shipboard sets, of which Fig. 26 is an example, cover a range of size, complexity, and performance nearly as great as that of land stations. Large naval vessels carry many different sets, each particularly designed

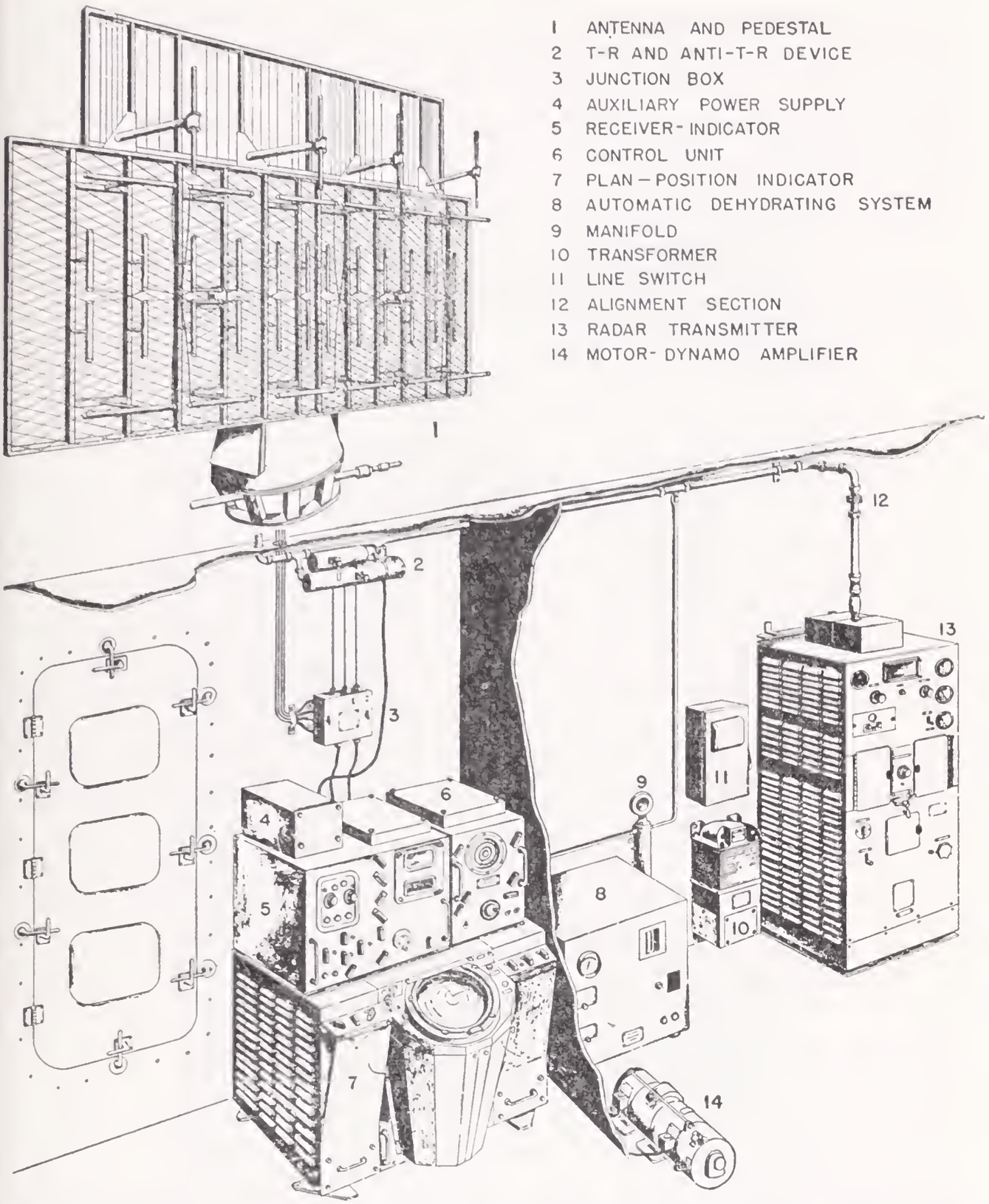


FIG. 26. Example of a shipboard set (the SC-3) used for search purposes. The set operates in the *P* band at a frequency of approximately 220 Mcps. The pulse duration is 5 μs, the repetition frequency is 60 cps, and the peak power is 200 kw. (*Courtesy of General Electric Company.*)

for a specific function and many of complex design. Remote plan-position indicators are used with these sets so that the radar indications may be observed at several different positions on a ship. Simpler radar equipment is carried by some merchant vessels.

Aircraft radar sets are necessarily fairly small and light in weight, though complicated circuit arrangements are often employed. Large

antennas cannot be accommodated on an aircraft, and therefore, high frequencies are necessary for adequate directivity. Maximum ranges tend to be less than those of land-based or shipboard equipment. The photograph of the AN/APS-4 set (Fig. 27) serves to illustrate the compact construction of aircraft sets.

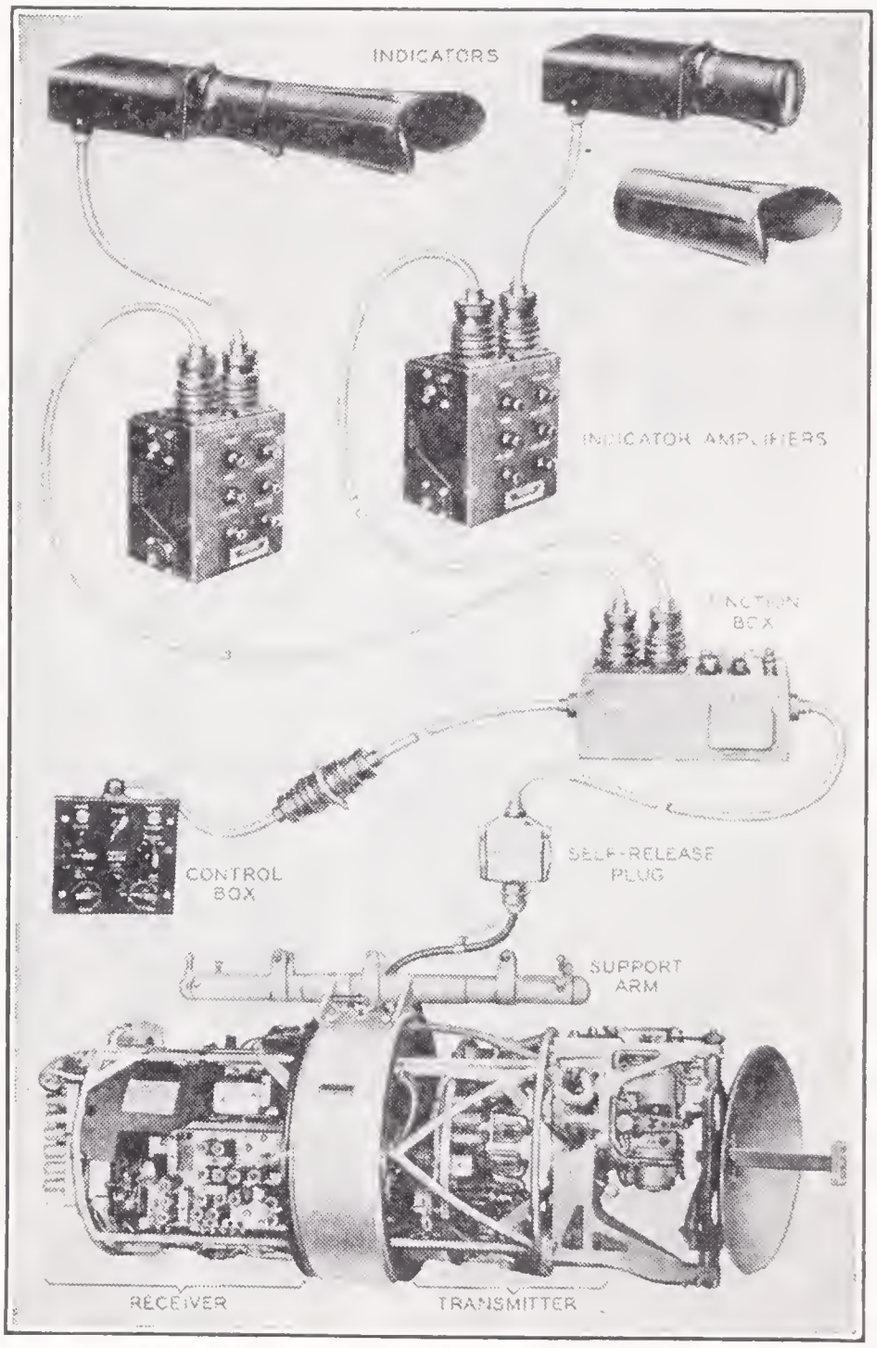


FIG. 27. Example of an aircraft set (the AN/APS-4) used for search purposes. The operating frequency is approximately 9,000 Mcps, the pulse duration is 0.6 μs, the pulse repetition frequency is adjustable to either 600 or 1,000 cps, and the peak power is 40 kw. (*Courtesy of Bell Telephone Laboratories.*)

A second classification method is based on the functions of radar equipment. There are, for example, the wartime functions of *search* (scanning air and sea for enemy craft) and *fire control* and the peacetime functions of air and sea *navigation* and *traffic control.* In some applications radar equipment having very special capabilities is required—such as the ability to measure very short ranges accurately, the ability to measure

the speed of moving objects, or the ability to detect weak echoes from moving objects despite a clutter of strong echoes from near-by stationary objects.

Search-type radar sets need not have exceptional accuracy or resolution but are expected to detect enemy craft at the greatest possible range. These sets may be further classified as *air search* or *sea search*. For air search, the lower radar frequencies are customarily employed to provide long ranges despite the weak aircraft echoes. For sea search and to detect low-flying aircraft, somewhat higher frequencies are better because energy of higher frequency can be directed closer to the surface of the ocean.

Fire-control applications of radar place special emphasis on the accuracy of range and angle indications. Resolution is necessary to accuracy, and therefore, fire-control sets typically operate at high frequencies and have short pulses. Antennas are large, pulse edges are steep, and receiver bandwidths are great. Many special circuits are included to permit rapid and accurate determination of range and coordinate angles. Land-based equipment is used with antiaircraft guns and harbor-defense guns. On shipboard, a different type of fire-control radar is required for each size of gun. In aircraft, accurate radar equipment may control not only gunfire but also the dropping of bombs.

Operation of fire-control equipment usually differs from that of search equipment in that a single object is *tracked*, and the radar set furnishes range and angle data for only this one object. *Angle tracking* consists in keeping the antenna pointed directly at the object at all times. Lobe switching or conical scan (see Art. 6) is used to permit accurate pointing of the antenna. *Range tracking* consists in adjusting a range step or other steep-edged waveform so that the steep edge continuously coincides with the front edge of the echo pulse. The tracking may be done manually by a range keeper, a pointer, or a trainer, or it may be done automatically. In *automatic tracking*, error signals derived from the echo pulses actuate *servomechanisms* (Chap. V) that cause the antenna and range step to follow the object coordinates faithfully. In some applications, *aided tracking* rather than complete automatic tracking is employed. The antenna is made to follow the target automatically in azimuth and elevation, but a range handwheel is used to control the rate at which the range step moves. The operator, therefore, makes corrections with the handwheel only when the rate of change of range is not constant.

Aircraft and shipboard sets used for navigation are in many respects similar to the corresponding types of sea-search sets. In general, there is more emphasis on simplicity and low cost for sets used in nonmilitary applications. Also, for close-in navigation, high resolution of near-by objects is important.

Traffic control of ships in harbors, and especially of aircraft near landing fields or flying along established airways, can be facilitated by large, complex radar sets at fixed points on the ground. Air-traffic-control sets observe the positions of all aircraft in their vicinity and transmit data and instructions to the pilots. Special sets of this classification may be used to guide aircraft in approach and landing operations during periods of bad visibility.

Effective use of radar in navigation and, especially, in the control of air traffic requires the coordination of radar devices with other types of equipment. Voice communication between aircraft and ground stations is essential, and a wide variety of other methods of data transmission, including television, have been used. Many devices are employed that are not exactly pulsed radar in the form described thus far, yet do have some of its features.

The *radio altimeter* is an example of such a radarlike device. The pulsed type of altimeter actually is a low-power radar system that directs its energy downward in a very broad beam. The surface of the earth below returns a strong echo, and the two-way travel time indicates the altitude of the aircraft relative to the ground immediately underneath. Another type of altimeter especially suitable for accurate measurement of low altitudes employs a frequency-modulated continuous-wave transmitter. The altitude determines the difference of frequency between transmitted and reflected waves, and the beat frequency generated is used as a measure of altitude.

The classifications *primary radar* and *secondary radar* are sometimes made. Primary radar refers to radar sets of the types already described in which the objects detected do not carry special equipment; the radar indications depend solely on natural reflections. Secondary radar refers to pulsed radio-location systems in which the object located carries special equipment for receiving pulses and, possibly, for transmitting reply pulses. Many of the radarlike devices employed in navigation and traffic control may be called secondary radar.

Radar beacons are an important example. The beacon includes a receiver and transmitter and transmits reply pulses whenever pulses from a radar set are received. The replies are transmitted essentially instantaneously and appear as especially strong echo pulses on the radar indicator. Beacon replies are distinguished from radar echoes in two ways: (1) The reply is usually transmitted on a different carrier frequency from that of the radar set. Thus, if the radar receiver is tuned to the frequency of its own transmitter, only radar echoes appear on the indicator screen. If the receiver is tuned instead to the beacon frequency, the radar echoes disappear and beacon replies are displayed. (2) The

beacon includes a coding unit that modifies the reply pulses in some special way. In one scheme (range coding), the beacon transmits a series of closely spaced long and short pulses. These pulses form a dot-dash pattern on a Type B or Type P indicator. The pattern identifies the beacon, and the shortest range pulse indicates the distance from radar set to beacon. The appearance of beacon replies on a Type P indicator may be seen in Fig. 28.

Radar beacons may be established at fixed positions and used in navigation. A radar-equipped ship or aircraft can then identify a beacon by

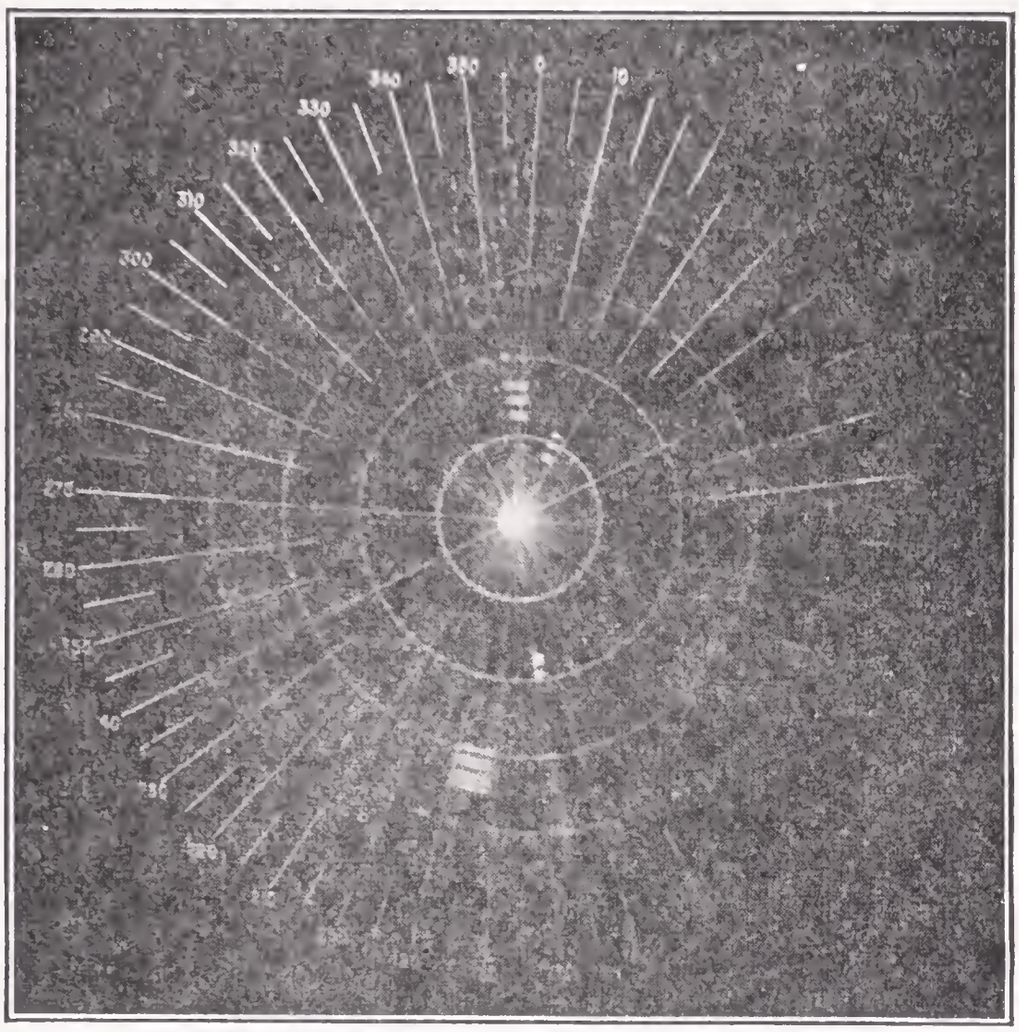


FIG. 28. Beacon replies on a Type P indicator. Four beacons may be identified in the picture.

its code and use the range and bearing of the beacon to determine its own location. In an alternate use, aircraft are equipped with lightweight beacons often called *transponders*. Ground radar stations used for traffic control may then locate and identify the aircraft more readily. The code of the transponder reply may be used to indicate the altitude of the aircraft, thus reducing the requirements made of the ground station. In the Second World War a system of identifying friend and foe used by the Allied Nations required each friendly ship or aircraft to carry a transponder adjusted to reply with the proper code.

A second form of secondary radar used in navigation is *loran* (*lo*ng *ra*nge *n*avigation). Ships and aircraft carry special receivers that measure very accurately the difference in time of arrival of pulses from a pair of transmitters. The transmitters are synchronized, and therefore the time difference is a measure of the difference of the distances between receiver and the two transmitters. The transmitter positions are known, and therefore the position of the receiver is known to lie on a curve drawn through all points of the specified difference of distances. This curve is a hyperbola. To determine position by loran, the time differences of the pulses from at least two pairs of transmitters are determined, and the intersections of the corresponding hyperbolas are determined. Because transmission of radar-frequency waves is limited by horizon effects, loran uses much lower frequencies—about 2 Mcps. Signals from each pair of transmitters may therefore be received over a large area.

9. Block Diagram of a Radar Set. The block diagram of Fig. 29 is intended to link the discussion of the general features of radar in preceding articles with the technical circuit details treated in subsequent chapters. Frequent reference to this diagram during the reading of later chapters may clarify the relation of each small part of a radar set to the complete system.

Figure 29 applies to an imaginary radar set of rather simple design. The antenna beam is assumed to be broad in elevation and narrow in azimuth, so that information needed for the Type P presentation is obtained when the antenna is rotated continuously by the azimuth motor. The antenna control permits stopping the motor and pointing the antenna in any desired direction. Thus the Type A indicator may be used to examine closely the echoes returned from this direction and make accurate range measurements using the range step.

The repetition frequency of the system of Fig. 29 is controlled by the master oscillator in the modulator block. This oscillator may generate almost any waveform. Once in every master-oscillator cycle, the pulse-forming circuits in the modulator generate a high-voltage pulse of closely rectangular shape and short duration. This pulse is used as the plate supply voltage for the transmitting oscillator, and thus brief pulses of radio-frequency energy are generated. A trigger pulse is also generated by the pulse-forming circuits and used to initiate a cycle of operations in the indicator. Waveforms of high-voltage pulse, transmitted pulse, and trigger pulse showing time relationships appear in Figs. 30*A*, *B*, and *C*.

The pulse from the transmitter is conducted along a coaxial transmission line or a waveguide to the antenna, from which it is radiated. The high-power pulse passes by the receiver input line, but only a very minute fraction of it travels to the receiver because of the transmit-receive device. Shown in Fig. 29 is an anti-t-r device, used in conjunction with the t-r

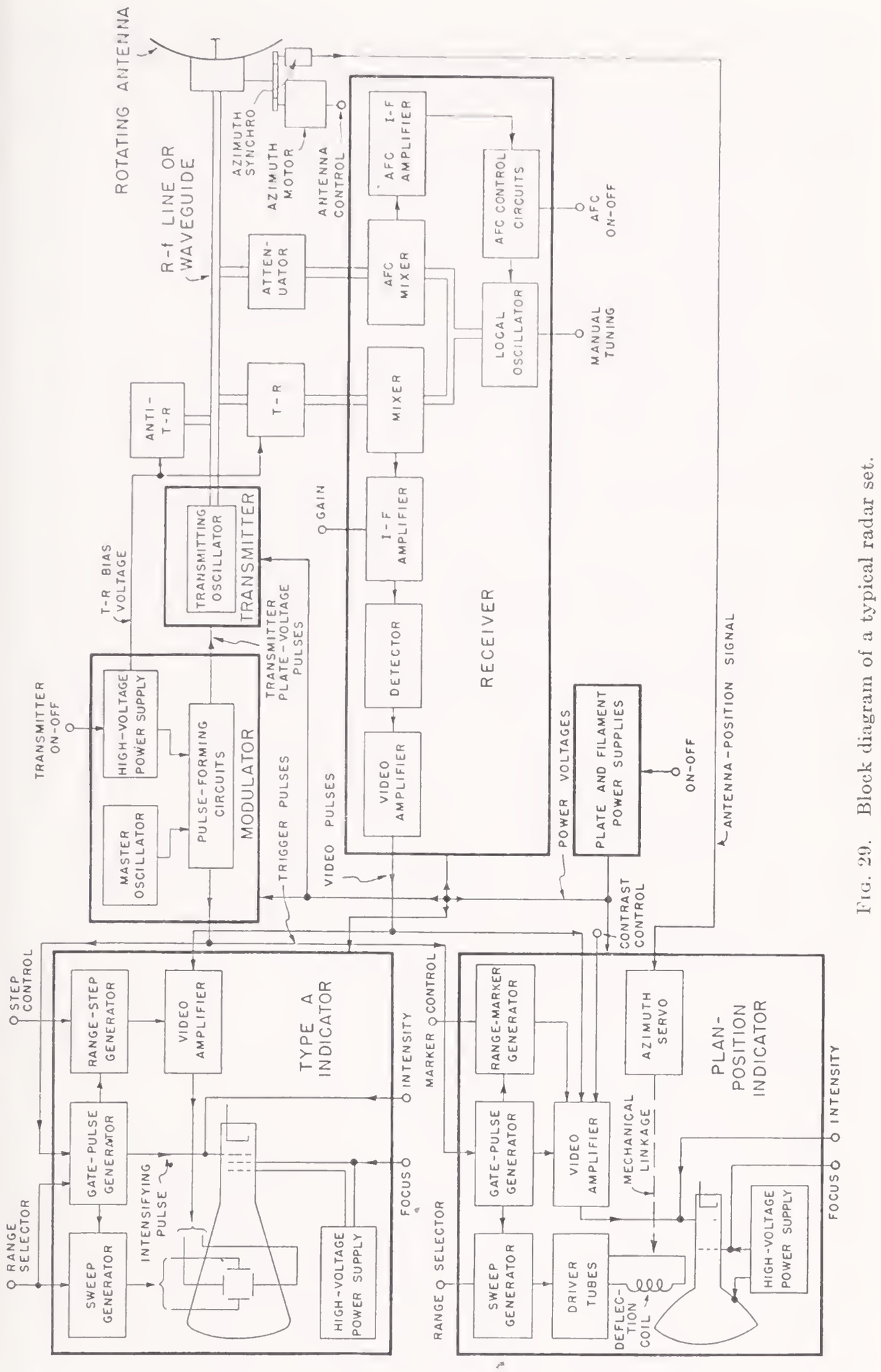


FIG. 29. Block diagram of a typical radar set.

device in many systems. Also shown is a connection from the negative terminal of a high-voltage supply to the t-r device. The anti-t-r device reduces the loss of echo power to the transmitter, and the t-r bias increases the receiver protection. The function of anti-t-r devices and t-r bias voltage is explained more fully in Arts. 10 and 11, Chap. XII.

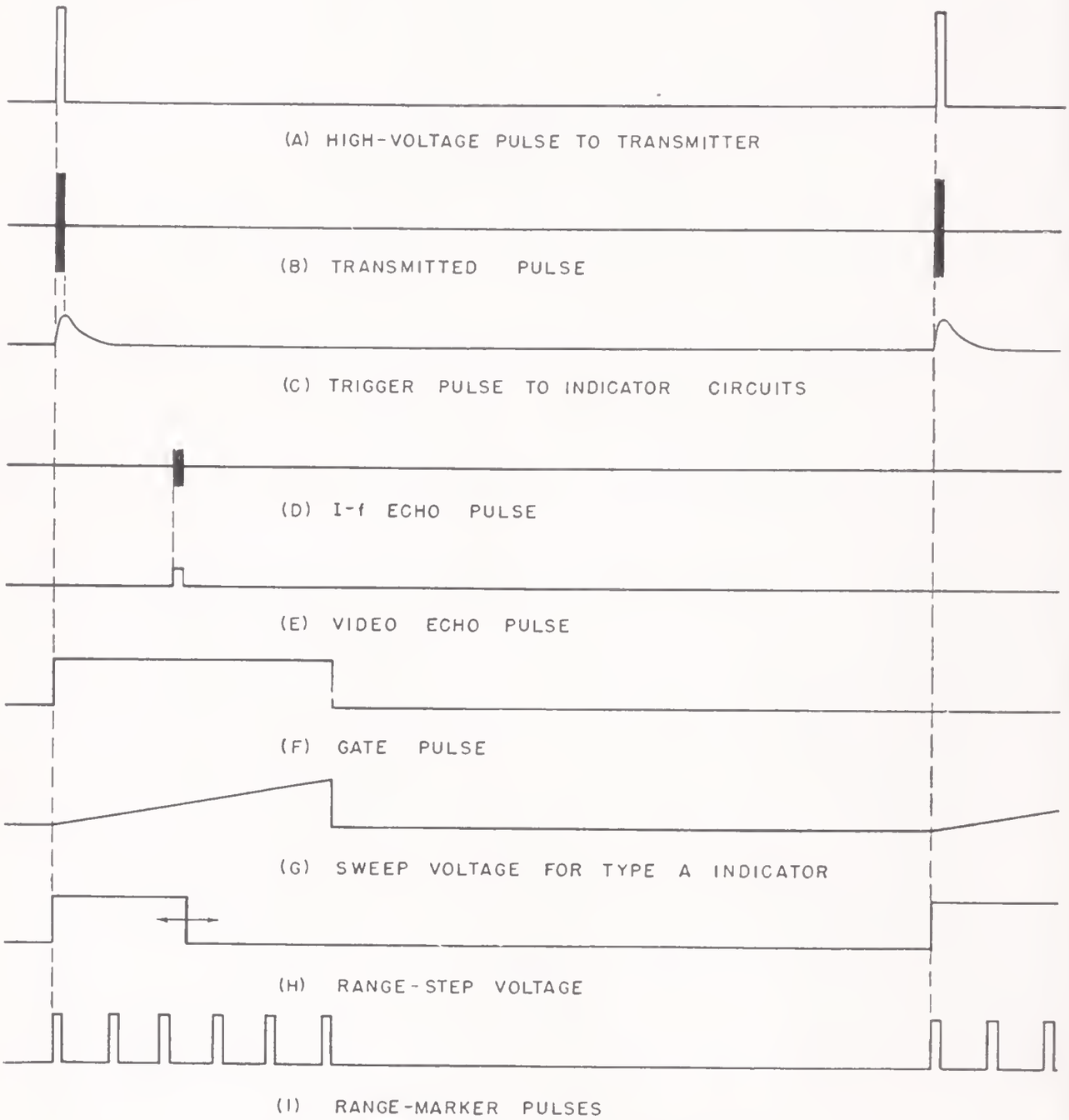

(A) HIGH-VOLTAGE PULSE TO TRANSMITTER
(B) TRANSMITTED PULSE
(C) TRIGGER PULSE TO INDICATOR CIRCUITS
(D) I-f ECHO PULSE
(E) VIDEO ECHO PULSE
(F) GATE PULSE
(G) SWEEP VOLTAGE FOR TYPE A INDICATOR
(H) RANGE-STEP VOLTAGE
(I) RANGE-MARKER PULSES

FIG. 30. Waveforms in radar set of Fig. 29.

Echo pulses returned to the antenna are conducted through the t-r device to the receiver. The echo pulse and local-oscillator signal are both applied to the mixer, and a beat frequency or difference frequency is generated. The local oscillator is adjusted so that the beat frequency is the intermediate frequency to which the i-f amplifier is tuned. The i-f amplifier comprises many stages and increases the voltages of the i-f echo pulses in the order of a million times. The detector rectifies the alter-

nating current of the i-f echo pulses and produces video pulses that follow the i-f pulse envelope. Simplified waveforms of i-f and video echo pulses are drawn in Figs. 30*D* and 30*E*.

The automatic-frequency-control, or afc, system in the right-hand part of the receiver block is provided in order that the receiver may be kept continuously in tune with the transmitter, despite effects of temperature changes, changes in transmitter loading as the antenna rotates, and so on. By means of an attenuator, a small fraction of the transmitted pulse in the main r-f line is supplied to the afc mixer. The local-oscillator voltage also is applied to this mixer. The afc and main mixers function in the same way, except that almost no echo power is supplied to the afc mixer —the afc system is operated by the transmitted pulses. The i-f pulses in the afc system are amplified and applied to control circuits that generate a direct voltage dependent on the frequency of the pulses. This voltage tunes the local oscillator. If the i-f pulses have the correct frequency, the direct voltage has a value that merely maintains the local-oscillator frequency. If the i-f pulses have too high a frequency, the direct voltage changes and shifts the local-oscillator frequency in the direction to reduce the intermediate frequency; too low an intermediate frequency causes the opposite shift of local-oscillator frequency.

Video pulses from the receiver are applied to both indicators. Consider first the Type A indicator. Video signals are applied to one pair of deflecting plates in the cathode-ray tube and cause vertical motion of the spot on the screen. The sweep voltage is applied to the horizontal-deflection plates to produce the Type A presentation. This voltage is obtained from a sweep-generating circuit that is controlled by a rectangular pulse from the gate-pulse generator. Waveforms of gate-pulse and sweep voltages appear in Figs. 30*F* and *G*. The leading edge of the gate pulse is timed by the trigger pulse from the modulator, and the duration is determined by the gate-pulse generator itself. The gate-pulse determines the start and end times of the sweep, and the sweep generator controls the rate of rise of the sweep voltage. The gate pulse is used also as an intensifying pulse. It is applied to the grid of the cathode-ray tube so that the spot will be formed on the screen only during the sweep time, thus eliminating signals that occur during the return trace of the spot. A third function of the gate pulse is to control the range-step generator. This circuit measures time from the leading edge of the gate pulse and forms a step of voltage, as shown in Fig. 30*H*, after an accurately controlled interval. The step voltage is added to the echo pulses in the video amplifier. The step can be adjusted by the step control until it coincides with an echo of interest, and the range of the object causing the echo read from calibrations on the step-control dial.

In the plan-position indicator, deflection of the electron beam of the

cathode-ray tube is produced by the magnetic field of the deflection coil. Video signals are applied to the grid of the cathode-ray tube so that they increase the intensity of the luminous spot. The spot is deflected in two ways: (1) It is moved from the center of the tube to the edge at a definite rate to indicate range. This motion is produced by current in the deflection coil controlled by a sweep generator. The sweep generator is similar to the one in a Type A indicator but employs driver tubes having a power capability of several watts to furnish the saw-tooth current wave to the coil. (2) The radial line caused by range sweep of the spot is rotated around the face of the tube as the antenna turns in azimuth. In Fig. 29, this rotation is accomplished mechanically by making the deflection coil revolve about the neck of the cathode-ray tube. An azimuth synchro (discussed in Chap. V) is geared to the antenna and generates voltages dependent on angular position. These voltages, applied to another synchro in the azimuth servomechanism, cause the servo motor to rotate the deflection coil in correspondence with the antenna. To make approximate range measurements possible with the Type P indicator unit, the range-marker circuit is provided. This circuit is controlled by the gate pulse and generates a series of evenly spaced narrow pulses, as shown in Fig. 30*I*. These pulses are combined with the video signals and appear on the indicator screen as a series of concentric circles indicating known values of range. Use of an intensifying pulse is essential with an intensity-modulated presentation to ensure that the luminous spot disappears between sweep intervals. Therefore, the gate pulse is combined with the other signals in the video amplifier.

Most of the functions accomplished by the circuits of Fig. 29 can be accomplished in a number of different ways, only one of which is presented in Fig. 29. Other possibilities are discussed in later chapters. Of general importance is the over-all timing system. With respect to timing, the circuit of Fig. 29 can be called a *master-oscillator* system, because a master oscillator controls the repetition rate. Some triode transmitters are self-pulsed; that is, the transmitter automatically operates in short bursts (Art. 12, Chap. X) and no modulator or master oscillator is needed. Other systems employ modulators in which a pulse is produced each time a spark occurs between a point on a rotating wheel and a fixed point (Art. 8, Chap. III). Systems of these types are sometimes called *self-synchronous*. In them a trigger pulse must be derived from the transmitted pulse to time the indicator circuits. The master-oscillator system has an advantage in flexibility. For example, in it the trigger pulse for the indicator circuits may be made to occur a definite time earlier than the transmitted pulse to make up for delays in the indicator circuits. On the other hand, the self-synchronous system is often simpler. Furthermore, erratically variable delays that occur in

some modulator circuits can affect the range accuracy of a master-oscillator system unless the indicator trigger is derived from the transmitted pulse rather than directly from the master-oscillator output.

The operating controls for the radar set in Fig. 29 are indicated in the figure by small circles. The operation and functions of these controls may be outlined as follows:

Power Controls. Power to all circuits except the modulator and transmitter plate-voltage supply is controlled by the main ON-OFF switch. An additional switch is provided for the transmitter high voltage in all but the lowest power sets, because the cathodes of the power tubes must be allowed to heat before the high voltage is applied. Often a time-delay relay is included to prevent closing the high-voltage circuit too soon.

Type A Indicator Controls. The *intensity* and *focus* controls permit adjustment of the indicator-tube voltages for an easily visible, sharply defined range trace. Too great intensity causes a glow to surround the trace and prevents sharp focus. Most sets provide several different range sweeps, and the *range selector* is used to choose a long-range sweep for observation of echoes from distant objects or a short-range sweep for detailed examination of near-by objects. The selector control adjusts both the gate-pulse generator and the sweep generator. The gate-pulse duration must always correspond to the maximum range of the sweep in use, and the sweep generator must produce a faster spot motion on the shorter range sweeps. The *step control*, as previously explained, determines the range at which the step occurs and permits accurate range measurement.

Plan-position-indicator Controls. The *range selector*, *focus*, and *intensity* controls function in the same manner as the corresponding Type A indicator controls. A different adjustment of intensity is used, however. With the receiver gain control set for minimum gain the intensity is adjusted for about the threshold of visibility of the range trace. When the gain is increased, echo pulses brighten the trace in spots and indicate the location of objects. If the receiver gain is made high, so that weak echoes produce visible indications, there is a possibility that strong echoes may cause too intense a spot that spreads out from its proper position and obscures weak echoes. To avoid this effect, the final video-amplifier stage includes a means of limiting the amplitude of strong signals and the *contrast control* adjusts the point at which limiting occurs. It serves as a maximum-intensity control and may be set for a high intensity to increase the contrast between strong-signal and weak-signal indications or low intensity to increase the visibility of weak-signal indications. The *marker control* may be simply an ON-OFF switch so that marker pulses can be made to appear only when wanted, or it may be a control to provide any desired intensity of marker pulse.

Receiver. In order for echo signals to be observed, the local oscillator of the receiver must be correctly tuned. Some types of afc circuits, when operating properly, will accomplish the tuning automatically as soon as the *afc switch* is set to ON. Other types require that the receiver first be tuned manually with afc off. If the afc switch is then turned on, the tuning is maintained automatically. When manual tuning is required, the focus and intensity of the indicators must first be adjusted and the receiver *gain* increased until either echo signals or noise pulses are observed. The *manual-tuning* control is then adjusted for maximum amplitude of the echo pulses. If the signals are strong, the gain must be decreased as the tuning progresses in order that the variation of signal strength with tuning may be observed.

Correct use of the gain control is important in obtaining maximum information from a radar display. If weak signals are to be detected, a high gain is required so that discernment of signals may be limited only by noise. Visibility of weak signals depends upon the combination of the setting of gain, intensity, and contrast controls, and considerable experience with a particular set is required in order to adjust the controls to best advantage. If angular resolution is wanted, the gain must be set as low as signal visibility permits. Otherwise, the echo indications spread out in azimuth because the edge of the antenna beam is not sharply defined. Generally, the various parts of a single range trace should be viewed with different gain settings—high gain makes the weak echoes at long range visible, and low gain helps to resolve the many different strong echoes at short range.

Components of two very different kinds make up the radar set of Fig. 29. The *low-frequency components* are electric circuits made of resistors, capacitors, inductors, vacuum tubes, motors, and so on, not unlike the many circuits widely used for power and communication. Such circuits as these appear in the modulator, indicators, receiver, and power supply and in the devices used to control the antenna azimuth motor. The *radio-frequency components,* on the other hand, are devices in which the ordinary circuit parameters do not appear separated from one another. These components include transmission lines, waveguides, antennas, and the special tubes employed in microwave radar. These tubes are distinguished by the important part the time of transit of electrons within the tubes plays in tube operation and by the resonant cavities built into the tubes to serve as tuning elements.

In accordance with this division of radar-set components, the remaining chapters of this book may be divided into two parts. Chapters II through VI deal with low-frequency components. Chapter II, "Timing Circuits," explains the basic principle of a class of circuits much used in radar, especially in indicators and modulators, and the next four chapters

make application of these principles. Chapters VII through XIII discuss the radio-frequency components, beginning with principles in "Radio-frequency Lines" and "Waveguides" and detailing the applications in the subsequent chapters. The final chapter, "Propagation," extends ideas about radio waves beyond the radar set itself and considers the propagation of energy from set to a reflecting object and back.

CHAPTER II

TIMING CIRCUITS

The majority of low-frequency circuits in a radar set serve a timing or control function. These circuits produce a variety of voltage waveforms such as saw-tooth waves, trapezoidal waves, rectangular pulses, and sharp peaks, and the tubes in these circuits are usually violently overdriven, the operating point moving from far below cutoff to the region of grid current. All these circuits are broadly known as *timing circuits*, though many of them may be more specifically classed as modulator circuits, indicator circuits, range-measuring circuits, or even receiver video circuits, depending upon their particular functions. The present chapter is concerned with the principles of all timing circuits and provides a foundation upon which discussions of special circuits in later chapters are built.

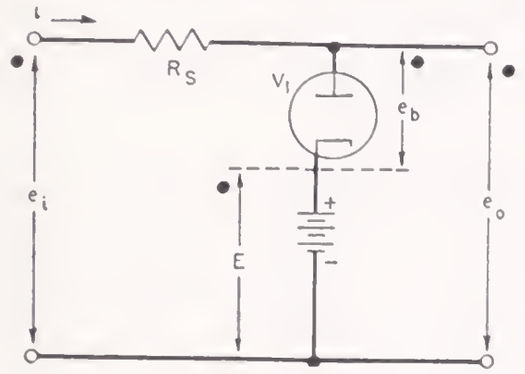
A. CLIPPING CIRCUITS

Because amplifier tubes in timing circuits usually operate with input signals and bias voltages that cause plate current to be completely cut off during a part of the cycle or grid current to flow during another part or both, the output waveform is seldom a reproduction of the input. Instead, certain portions of the input wave are completely missing from the output. Tubes in such circuits are said to be *overdriven*, and the waveforms produced are said to be *clipped*. Diodes also are useful for clipping because they conduct current in only one direction.

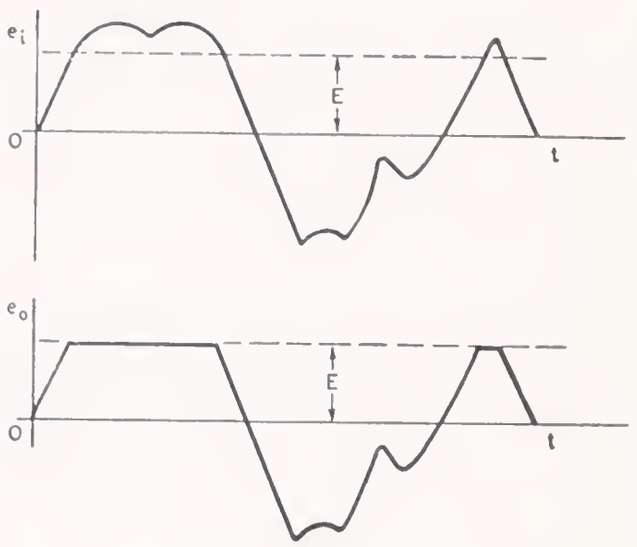
1. Clipping with Diodes. Figure 1A is the circuit diagram of a simple diode clipper. The input voltage is e_i, and the output voltage e_o. The load connected to the output terminals is assumed to have a very high impedance, so that the output current is negligible. The series resistance R_s is necessary to make possible an output voltage different from the input voltage. In some circuits a physical series resistor is used; in others R_s is the internal resistance of the source of input voltage.

If the input voltage e_i is less than the battery voltage E at any instant, the net voltage in the circuit tends to send current from cathode to plate through the tube. Thus, the tube behaves as an open circuit, no voltage drop occurs across R_s, and the output voltage e_o equals e_i. If e_i rises above E, the net voltage sends current from plate to cathode through the diode, and the diode behaves as a short circuit connecting the top output terminal to the positive battery terminal. The output voltage equals E, and the difference between e_i and E appears as a voltage drop across R_s.

If any input voltage wave, such as e_i in Fig. 1*B*, is applied to this circuit, the output voltage follows the input whenever both are less than E but is limited to the voltage E whenever the input exceeds this amount. In this way the positive peaks of the input wave are clipped at a voltage level set by the battery.



(A) CIRCUIT DIAGRAM



(B) WAVEFORMS

FIG. 1. Diode clipper.

So far the diode has been assumed to behave as a short circuit when conducting. This assumption is justified only if R_s is very large compared with the resistance of the diode. The *static, d-c,* or *average* plate resistance of the diode is defined as

$$r_b = \frac{e_b}{i_b} \tag{1}$$

where

e_b is the plate voltage of the diode,
i_b is the plate current of the diode.

This resistance is nonlinear; that is, its value changes as i_b and e_b change. As an approximation, r_b is often assumed to have a constant value whenever e_b is positive.

To see the effect of this resistance, suppose that R_s in Fig. 1 is 4,500 ohms (4.5 K), and that r_b for the diode is 500 ohms (0.5 K). The current flowing when e_i is greater than E is then

$$i = \frac{e_i - E}{4.5 + 0.5} = \frac{e_i - E}{5} \qquad \text{ma}$$

where e_i and E are in volts,[1] and the output voltage is

[1] In calculating the small currents encountered in electronic circuits, convenient units to use are *volts* and *kilo-ohms,* the currents being found in *milliamperes.* This set of units and other similar combinations are used frequently in numerical examples in this text.

$$e_o = E + e_b = E + ir_b = E + \frac{e_i - E}{10} \qquad (2)$$

Equation 2 shows that a fraction of the excess voltage of e_i over E—in this example $\frac{1}{10}$—is added to E in the output voltage. If effective clipping is to be had, R_s must be very much larger than r_b. In the example, if R_s were changed to 0.5 megohm, only 1/1,000 of $e_i - E$ would appear in the output and clipping would occur almost exactly at the voltage E.

The method of indicating directions of currents and voltages on diagrams should be noticed. In Fig. 1*A* the arrow near i does not indicate that current *must* flow in the direction of the arrow; rather, it means that *if* current is in that direction, it is called *positive;* if in the reverse direction, it is called negative. Because of the diode, current *can* flow only in the direction of the arrow. Therefore, i must always be positive or zero in this circuit. Similarly, the voltage e_i is called positive if the potential at the end of the arrow marked with a dot is positive relative to the potential at the other end of the arrow. Voltage e_i is called negative if the dotted end is at a negative voltage relative to the other end. The dots by the e_o, e_b, and E arrows have a similar meaning. The plus and minus signs by the battery indicate the actual battery polarity.

2. Clipping with Amplifiers. The circuit diagram of a simple overdriven amplifier appears in Fig. 2*A*. Though a triode tube is shown, the principle of operation is the same if a tetrode or a pentode is employed. The series grid resistor R_s may be a physical resistor or the internal resistance of the source of signal.

The notation for tube currents and voltages is indicated in Fig. 2*A*. The symbols are defined as follows:

e_b is the plate voltage, measured with respect to cathode.
e_c is the grid voltage, measured with respect to cathode.
i_b is the plate current.
i_c is the grid current.

All these symbols stand for *total instantaneous values,* that is, actual voltage or current at any instant, not merely an a-c component. Reference directions are chosen so that e_b (or e_c) is a positive number whenever the plate (or grid) is actually positive with respect to the cathode and i_b (or i_c) is a positive number when conventional current flows in the direction of the arrows. These standard symbols are used throughout this text.

The numerical values indicated in Fig. 2*A* are typical and are to be used in an illustrative example. For the example, assume the input voltage e_i to be a sine wave of 100-volt peak value, as shown in Fig. 2*B*.

The grid-cathode path of the tube and resistor R_s form a diode clipper like that of Fig. 2*A*, but with $E = 0$. Therefore the positive peaks are

removed from the wave of e_i, while the negative peaks are unaltered, and the voltage actually applied to the grid is the wave of e_c shown in Fig. 2*B*. The elimination of the positive half of the grid-voltage wave because of grid current is called *grid-circuit clipping*.

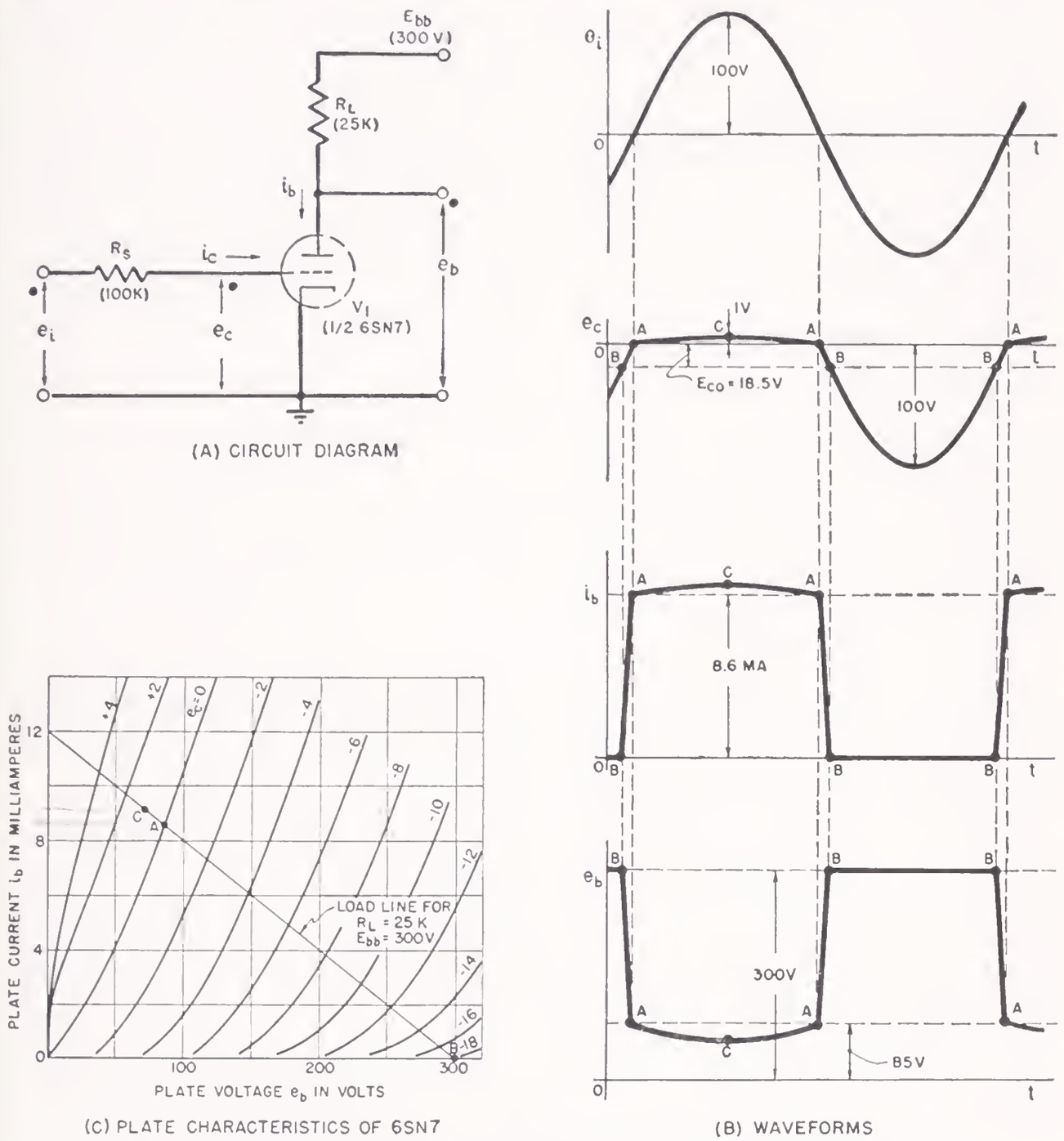


FIG. 2. Amplifier in which grid-circuit and plate-circuit clipping occur.

The amount by which the grid voltage rises above zero in spite of the clipping may be determined in the same way as the rise of e_o above E was determined for the diode clipper. The *static*, *d-c*, or *average* grid resistance is defined as

$$r_c = \frac{e_c}{i_c} \tag{3}$$

that is, the static grid resistance is the ratio of *total* grid voltage to *total* grid current. Like r_b of a diode, this resistance is nonlinear. Its value

depends upon both the grid voltage and the plate voltage. However, if the grid voltage of a small receiving-type tube is positive, and if the plate voltage (or screen-grid voltage of a tetrode or pentode) is at least three or four times as large as the grid voltage, then a useful approximation is obtained by considering r_c to be constant. For one triode section of a Type 6SN7 tube, r_c is approximately 1,000 ohms; for other small receiving-type tubes, the value of r_c is not greatly different.

Application of the approximation $r_c = 1{,}000$ ohms to the example of Fig. 2 shows that, during the positive half cycle of the e_i wave, the input voltage is applied across resistances of 100 K and 1 K in series. The part of the voltage that appears across the 1-K resistor is the grid voltage and, by the voltage-divider relation, has a peak value of approximately 1 volt.

The grid-voltage waveform, which is now known, may be used with the plate characteristics and load line of Fig. 2*C* to determine plate-current and plate-voltage waveforms. Any load that may be connected to the output terminals is assumed to be of such high impedance that its current is negligible. Therefore, resistor R_L determines the slope of the load line. Values of e_c at various points of the waveform may be referred to the load line, and corresponding values of i_b and e_b determined. The essential features of the output wave may be found by use of only two or three points as follows:

Point A. $e_c = 0$. Values of e_b and i_b for point A in Fig. 2*C* determine points A in the waveforms of e_b and i_b of Fig. 2*B*. The plate voltage is much less than the plate supply voltage because the plate current causes a large voltage drop across the load resistor.

Point B. $e_c = -E_{co}$. The negative grid voltage that is just sufficient to cut off plate current is called the cutoff grid voltage, and its magnitude is denoted by E_{co}. In Fig. 2, E_{co} equals 18.5 volts. A convenient method for determining the cutoff portion of the output waves is to draw a line on the e_c wave at the cutoff voltage and to project from intersections with the waveform, as indicated in the figure.

The value of the cutoff voltage is best determined from the plate characteristics, as done here. If characteristics are not available, the following *approximate* relation may be used for *triode* tubes:

$$E_{co} = \frac{e_b \text{ at time of cutoff}}{\mu} \tag{4}$$

where μ is the amplification factor of the tube. For the Type 6SN7 tube, $\mu = 20$ and, in the example given, e_b at the time of cutoff is the plate supply voltage, 300 volts.[1] Therefore, Eq. 4 gives a cutoff voltage of -15 volts which may be compared with the actual value, -18.5 volts.

[1] In circuits containing inductances or capacitances, the value of e_b at the time of cutoff is *not* necessarily the plate supply voltage.

The changes of i_b and e_b between points A and B are so rapid that it is sufficiently accurate to draw straight lines connecting these points. Thus the entire plate waveforms are determined except for the portions between points A. These remaining portions are very nearly horizontal lines. Their actual shapes may be determined by locating an additional point on the load line.

Point C. e_c = maximum positive value (+1 volt in the example). This point indicates the maximum value of plate current and minimum value of plate voltage and determines the deviation from a straight line of the portions of the waveforms between points A.

The flattening of the top of the e_b waveform is the result of the grid voltage going below the cutoff value and is known as *plate-current-cutoff*

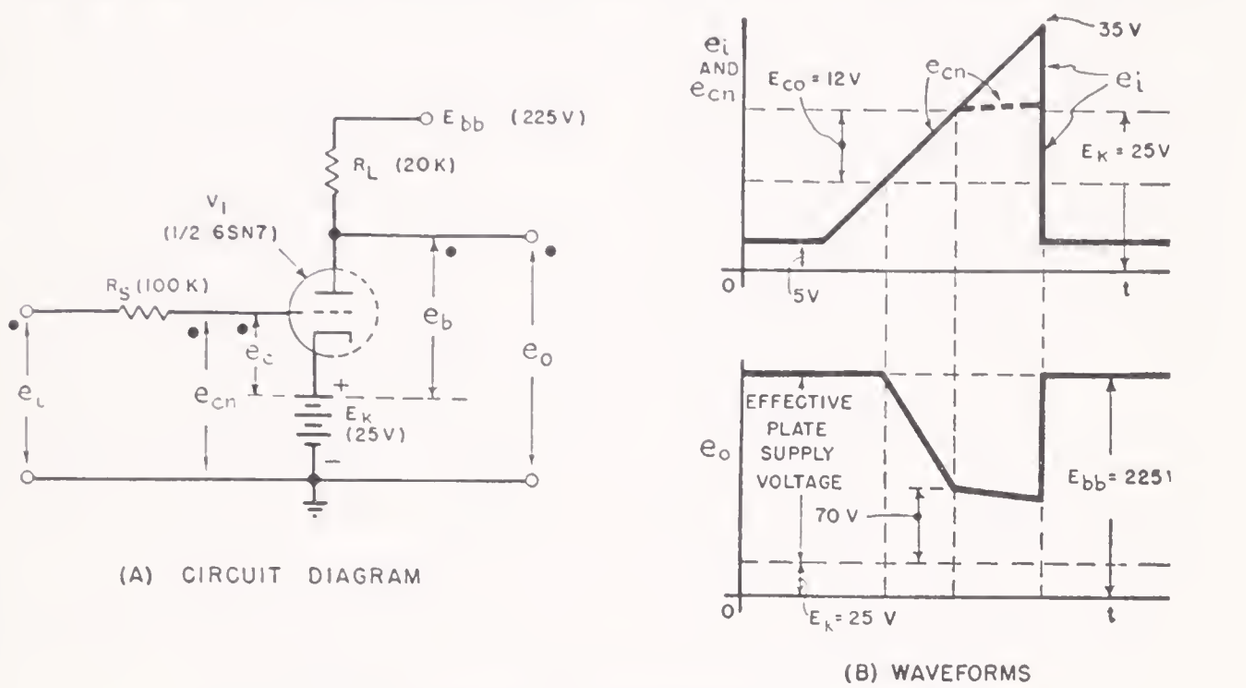


FIG. 3. Second example of clipping with an amplifier.

clipping. The combination of grid-circuit clipping (to flatten the *top* of the *grid-voltage* wave and *bottom* of the *plate-voltage* wave) and plate-current-cutoff clipping (to flatten the *top* of the *plate-voltage* wave) results in converting a sinusoidal input wave into a nearly square output wave. By use of a second amplifier stage of the same type, an almost perfect square wave can be obtained.

A second example of clipping is afforded by the circuit of Fig. 3*A* to which the saw-tooth wave e_i of Fig. 3*B* is applied. This example shows that a bias voltage in the cathode circuit changes the voltage level on the grid-*to-ground* input wave at which grid-circuit clipping occurs, yet the grid-*to-cathode* voltage for clipping is still zero. (The waveform of e_c is the same as the waveform drawn for e_{cn}, but with the zero axis moved to the $+E_k$ line.)

The equivalent circuit of Fig. 4*A* is sometimes useful in analyzing a clipping amplifier. The switch S_1 is open when the grid voltage is nega-

tive and closed when grid voltage is positive. The plate-circuit switch S_2 is open when the grid voltage is below cutoff and closed otherwise. The resistance r_p and voltage E_o are chosen to simulate the tube characteristics during plate-current flow.

To see how the plate circuit of a tube is represented by a resistance and a battery, assume that in a particular circuit $E_o = 20$ volts and $r_p = 7.5$ K. The relation between e_b and i_b *in the equivalent circuit of Fig.* 4*A* when switch S_2 is closed is then

$$e_b = 20 + 7.5i_b \tag{5}$$

where i_b is in milliamperes and e_b in volts. A plot of Eq. 5 is the dashed straight line in Fig. 4*B* which intersects the e_b axis at 20 volts and has a slope of 1/7.5 ma per volt. The values chosen for E_o and r_p make this

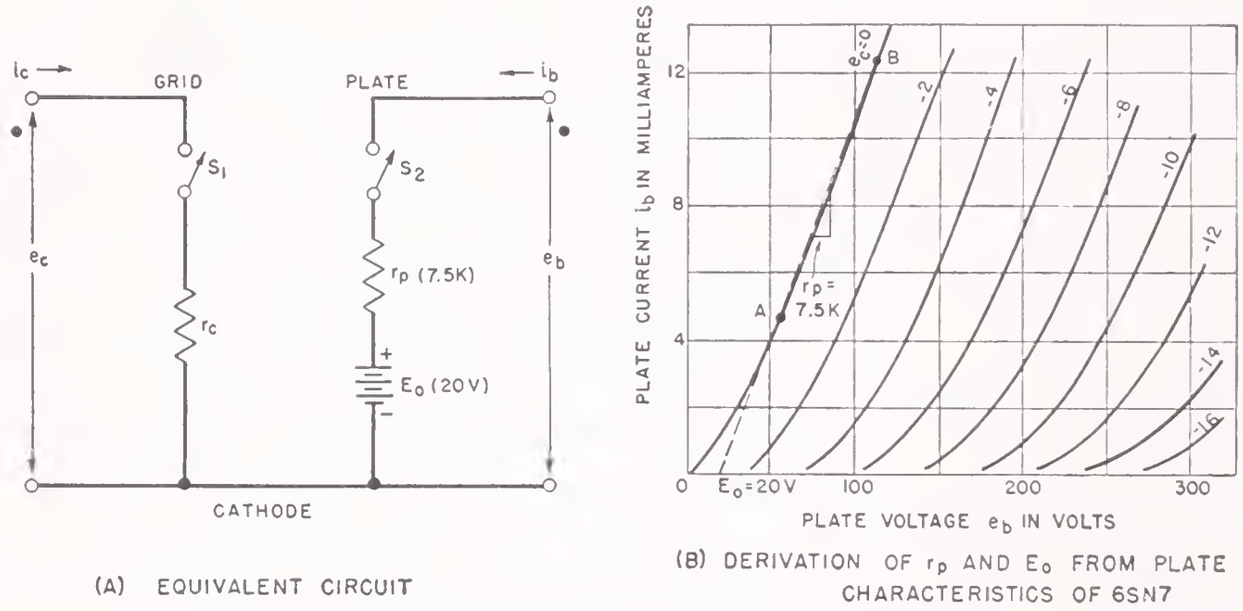


FIG. 4. Equivalent circuit for overdriven tubes.

line very nearly coincide with the $e_c = 0$ line of a Type 6SN7 triode over the range between points A and B in the figure. Therefore the equivalent circuit using these values of r_p and E_o simulates an overdriven 6SN7, provided that only the following two operating conditions of the tube are important: (1) plate current cutoff (switch S_2 open) and (2) plate current flowing (switch S_2 closed) with $e_c = 0$ and e_b between the values corresponding to points A and B.

Resistance r_p is essentially the reciprocal of the slope of the plate-family curve approximated by the straight-line characteristic of the resistor and battery. Thus r_p is the usual *dynamic, a-c,* or *incremental* plate resistance defined by

$$r_p = \left.\frac{de_b}{di_b}\right|_{e_c=\text{const}} \tag{6}$$

The value of r_p varies from point to point in the plate family, but within a reasonably small region of the characteristics a good approximation is

had by assuming r_p to be constant and E_o to change when a shift from one constant-e_c curve to another is made. The incremental plate resistance r_p must be distinguished carefully from the static plate resistance $r_b = e_b/i_b$. The incremental resistance pertains to the slope of the plate curves, while the static resistance has to do with the slope of lines from the origin of the plate family to points on the curves. The static resistance is seldom useful because its value changes greatly when the operating point is changed.

Sometimes the series grid resistance in circuits like those of Figs. 2*A* and 3*A* is very small, so that little grid-circuit clipping occurs, large grid currents flow, and the grid voltage reaches large positive values. The bottom of the e_b wave is still flattened, however, because of nonlinearity in the plate circuit. No matter what voltage is applied to the grid, the plate voltage can never be reduced entirely to zero, for at least a small positive plate voltage is needed to produce the plate current that causes the voltage drop in the load resistor. Therefore, the plate-voltage waveform flattens along a line a little above the zero axis—the higher the value of R_L, the more nearly the plate voltage approaches zero. Clipping of this kind is sometimes said to be due to *plate-current saturation*. Plate current is limited not by emission of the cathode but by the plate supply voltage and load resistor—no matter how low the resistance of the tube is made, the plate current cannot exceed E_{bb}/R_L.

3. Clipping with Cathode Followers. The range of grid-voltage variation within which linear operation occurs is much wider for a cathode follower than for an amplifier. Nevertheless, signal and bias voltages applied to cathode followers in radar circuits are frequently such that clipping occurs. As an example, consider the circuit of Fig. 5*A* and the 300-volt (peak value) input sine wave shown in Fig. 5*B*. Resistor R_s may be either the internal resistance of the source of signal or a physical resistor. Because the grid-to-ground voltage is not the same as the grid-to-cathode voltage, the symbol e_{cn} is used for the grid-to-ground voltage; e_c is reserved for the grid-to-cathode value.

Assume that no appreciable current is taken from the output terminals. The load line for the cathode follower with $R_k = 25$ K may then be drawn as in Fig. 5*C*. Note that the load line is constructed exactly as for an amplifier—the current that flows through the tube and load resistor in series, for a given value of grid-to-cathode voltage, is not affected by the order of connection of the tube and resistor.

The use of the load line with a cathode follower is not so simple as with an amplifier because the grid-to-cathode voltage is not directly known; e_c is the input voltage minus the unknown output (cathode) voltage. The load line may, nevertheless, be used to plot the output-input characteristic of Fig. 5*D*, and from this curve values of e_k for given values of e_i

may be read.[1] To obtain the output-input characteristic, assume values of e_c, and calculate and plot corresponding values of e_k and e_i. For example, if $e_c = 0$, the point of operation is point A on the load line and $i_b = 8.6$ ma. Therefore $e_k = 25 \times 8.6 = 215$ volts.[2] Because $e_c = 0$ at

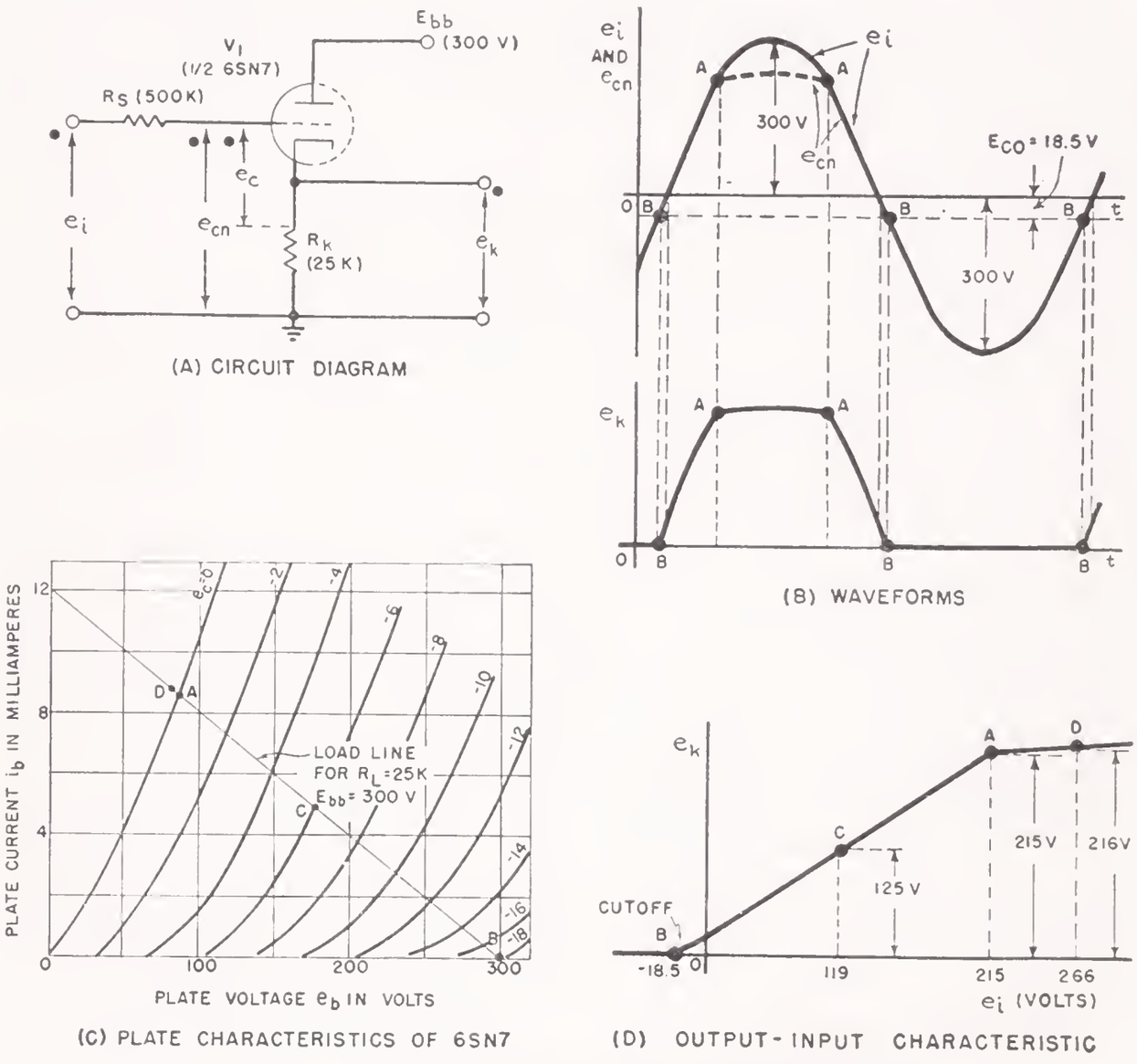


FIG. 5. Cathode follower in which clipping occurs.

this point, e_k and e_i are each 215 volts also, and point A may be plotted as in Fig. 5*D*. If e_c is reduced to approximately -18.5 volts or beyond, plate current is cut off, $e_k = 0$, and $e_i = e_c$. Thus the cutoff point B is located, and the curve to the left of point B in Fig. 5*D* lies along the horizontal axis. Points between A and B, such as point C, may be

[1] If e_k is to be found for only one or two given values of e_i, a shorter method is a trial-and-error calculation. Assume values of e_c, determine i_b from the load line, and calculate e_k and e_i. Repeat until the given value of e_i is obtained.

[2] In most of the calculations in this text, all the significant figures obtainable from a 10-in. slide rule are recorded in order to make the calculations easier to follow. The number of significant figures is *not* an indication of the accuracy of the result. Because of variations in tubes and other circuit elements, and because of the approximations made in the calculations, errors of 10 to 25 per cent may be expected frequently.

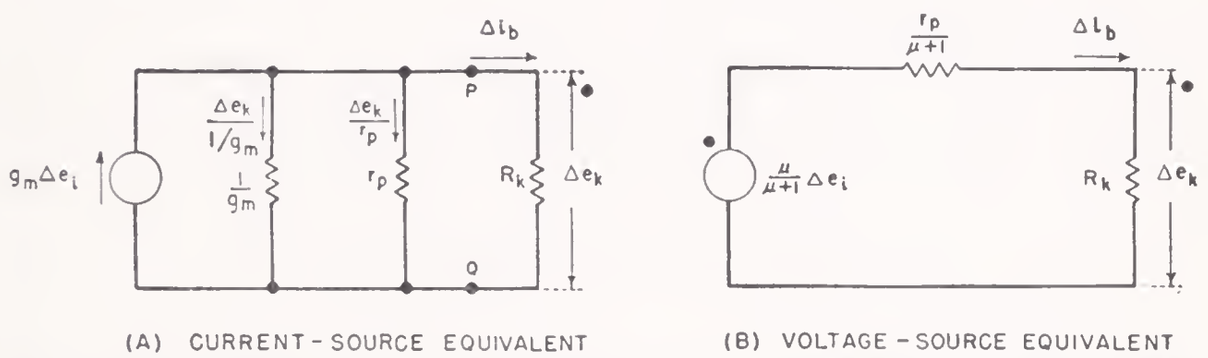
located similarly. Points beyond A correspond to positive values of e_c; therefore, grid current must be considered in locating them. For example, if $e_c = 0.1$ volt (point D in Fig. 5C),[1] $i_b = 8.65$ ma and $e_k = 216$ volts. If $r_c = 1$ K, $i_c = 0.1$ ma and

$$e_i = i_c R_s + e_c + e_k = 0.1 \times 500 + 0.1 + 216 = 266 \text{ volts}$$

Point D is plotted in Fig. 5D.

The waveform of e_k in Fig. 5B is obtained point by point by entering the characteristic of Fig. 5D with values from the wave of e_i. The same clipping effects produced in overdriven amplifiers may be observed in the waveform of e_k. Note, however, that polarity inversion does not occur in a cathode follower, and therefore grid-circuit clipping affects the positive portion of the wave, and plate-current cutoff affects the negative portion. Observe that cutoff occurs at the same point on the input wave for a cathode follower as for an amplifier, because for this point e_k is zero for the follower. Observe, also, that grid-circuit clipping begins when the *grid-to-cathode* voltage (not the input voltage) rises to zero. Because the cathode voltage is positive, grid current does not flow in the circuit of Fig. 5A until the input (grid-to-ground) voltage rises to +215 volts.

The cathode-follower characteristic of Fig. 5D is very nearly linear in the region between points A and B. In fact, for many purposes sufficient accuracy is obtained by locating points A and B and drawing a straight line between them. The slope of the output-input characteristic in the linear region is the voltage amplification, or ratio of a *change* in output voltage to the corresponding *change* in input voltage. The amplification is always less than unity because a change of output voltage e_k requires a change of e_c to produce it, and the change of the input voltage is the sum of the changes of e_k and e_c.



(A) CURRENT-SOURCE EQUIVALENT (B) VOLTAGE-SOURCE EQUIVALENT

FIG. 6. Equivalent circuits for cathode follower operating within its linear range.

The value of the voltage amplification may be found from either of the equivalent circuits of Fig. 6. The current-source equivalent circuit (Fig.

[1] Use of the load line in this calculation requires that the grid current be much less than the plate current in order that the effect of grid current in the cathode resistor may be neglected. In the example, $i_c = 0.1$ ma, and $i_b = 8.65$ ma.

6*A*) is obtained directly from the equation of linear operation of an amplifier-type tube:[1]

$$\Delta i_b = g_m \Delta e_c + \frac{1}{r_p} \Delta e_b \qquad \text{or} \qquad i_p = g_m e_g + \frac{1}{r_p} e_p \tag{7}$$

where Δe_c or e_g is a change in grid voltage, Δe_b or e_p is a change in plate voltage, and Δi_b or i_p is the corresponding change in plate current. Replacing Δe_c and Δe_b by their equals $\Delta e_i - \Delta e_k$ and $-\Delta e_k$ results in

$$\Delta i_b = g_m \Delta e_i - \frac{\Delta e_k}{1/g_m} - \frac{\Delta e_k}{r_p} \tag{8}$$

Equation 8 states that the output current Δi_b is the current $g_m \Delta e_i$ produced in a generator controlled by the input voltage, reduced by the currents in resistors of values $1/g_m$ and r_p. Therefore, an equivalent circuit for a cathode follower is the one shown in Fig. 6*A*.

The voltage-source equivalent circuit of Fig. 6*B* may be obtained from Fig. 6*A* by means of Thévenin's theorem. According to this theorem, the portion of the circuit to the left of points P and Q may be replaced by a resistance equal to the resistance between P and Q in the part replaced (measured with all voltage sources short-circuited and all current sources open-circuited) in series with a voltage source equal to the open-circuit voltage between P and Q. The resistance is $1/g_m$ in parallel with r_p, which equals $r_p/(\mu + 1)$. The open-circuit voltage is

$$g_m \Delta e_i \frac{r_p}{\mu + 1} = \frac{\mu}{\mu + 1} \Delta e_i$$

and thus the circuit of Fig. 6*B* is obtained. Both circuits in Fig. 6 apply only in the linear operating range of the tube, that is, between points A and B on the e_k, e_i characteristic.[2]

Either of these circuits may be used to find the voltage amplification A of a cathode follower. For the 6SN7 in the example of Fig. 5, $\mu = 20$ and $r_p = 7.5$ K. Therefore, according to the circuit of Fig. 6*B*,

$$\Delta e_k = \frac{20}{20 + 1} \times \frac{25}{25 + (7.5/20 + 1)} \times \Delta e_i$$

or

$$A = \frac{\Delta e_k}{\Delta e_i} = 0.94 \tag{9}$$

[1] This equation may be applied in a similar fashion to determine equivalent circuit for amplifiers and for many special circuits, such as single-tube phase inverters. See Art. 7, Chap. IV.

[2] The circuits may be applied to the region above point A, provided e_i is replaced by e_{cn} and provided grid current is very small compared with plate current.

Data from Fig. 5*D* show the average amplification over the entire range between points *A* and *B* to be

$$\left.\frac{\Delta e_k}{\Delta e_i}\right|_{\text{avg}} = \frac{215}{215 + 18} = 0.925$$

This average value is a little less than the value given by Eq. 9 because the tube is not exactly linear over the whole range, and the value in Eq. 9 is for an operating point near *A* where the e_k, e_i curve has its steepest slope.

Note that amplification applies to *changes* of voltage, not to actual values of voltage. The statement that a cathode follower has an amplification less than 1 does not mean that the output voltage e_k is less than the input voltage e_i. Rather, e_k is greater than e_i as long as the grid-to-cathode voltage is negative.

According to the equivalent circuit of Fig. 6*B*, a cathode follower may be considered as a voltage source having an internal resistance $r_p/(\mu + 1)$. This resistance is less by the factor $1/(\mu + 1)$ than the plate resistance of an amplifier tube. For the 6SN7 it is 7,500/21 = 357 ohms. Of course the output resistance of the circuit of Fig. 5*A* is slightly less, because by Thévenin's theorem it is the internal resistance of the tube in parallel with the 25-K load resistor, or 352 ohms.

The physical meaning of the low internal resistance of a cathode follower is that the output voltage is very nearly independent of the current taken from the output terminals. If, for example, the follower is operating at point *C* with an output voltage of 125 volts and an additional load of 10 ma is connected to the output terminals, the decrease of output voltage is only $0.010 \times 352 = 3.52$ volts. Notice particularly that the internal impedance is low only in the range between points *A* and *B*. If the cathode follower is cut off by a large negative input voltage, the tube itself becomes an open circuit, and the output impedance of the complete circuit in Fig. 5*A* is that of the cathode resistor, 25 K.

The input resistance of a cathode follower is sometimes of interest. If the grid-to-cathode voltage is positive, as at point *D* in Fig. 5, the input resistance is rather low and causes grid-circuit clipping. However, the input resistance is considerably higher than r_c, and therefore grid-circuit clipping with a cathode follower is less nearly perfect than that obtained with an amplifier.

An equivalent input circuit is drawn in Fig. 7. The switch *S* is included because grid current flows only when e_i exceeds the voltage *E*. To understand why the battery and the resistor appear in Fig. 7, consider the example of Fig. 5 in which, for operation at point *A* of Fig. 5*D*, $e_c = 0$ and $e_i = e_{cn} = 215$ volts. Grid current is zero for $e_{cn} = 215$ volts and increases nearly in proportion to the excess of e_{cn} over 215 volts when e_{cn} is

increased. This characteristic is the reason for representing the grid circuit as a 215-volt battery in series with a resistance.

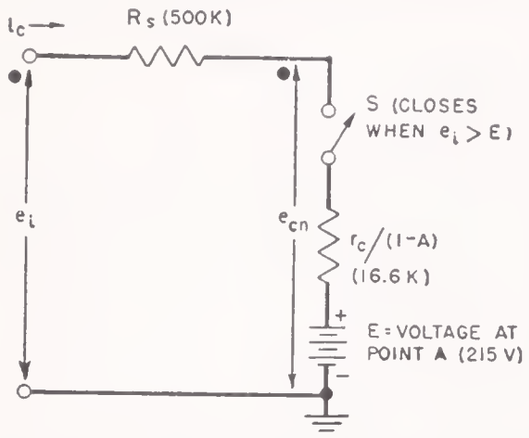


FIG. 7. Equivalent input circuit for cathode follower. Numerical values apply to circuit of Fig. 5A.

As e_{cn} increases above 215 volts, e_k in Fig. 5A increases, but the battery voltage $E = 215$ volts in Fig. 7 remains constant. Thus the voltage e_c across r_c in Fig. 5A is less than the voltage $e_{cn} - E$ across the resistor in Fig. 7. In order that the current in the equivalent grid circuit equal the actual grid current, the resistor in Fig. 7 must therefore be larger than the grid-to-cathode resistance. The magnitude of the larger resistance is obtained as follows: Let $e_{cn} = E + \Delta e_{cn}$; that is, let Δe_{cn} be the excess of e_{cn} over E. Then $e_k = E + A\Delta e_{cn}$, where A is the amplification of the follower in the region of grid-current flow. The grid voltage e_c is $e_{cn} - e_k = \Delta e_{cn}(1 - A)$, and therefore

$$i_c = \frac{e_c}{r_c} = \frac{\Delta e_{cn}(1 - A)}{r_c} = \frac{e_{cn} - E}{r_c/(1 - A)} \tag{10}$$

The resistance used in Fig. 7 is $r_c/(1 - A)$ in order that the current in the equivalent circuit may equal the actual grid current given by Eq. 10. In the example of Fig. 5, grid current has little effect upon amplification, and therefore A is nearly 0.94, the value obtained from the linear equivalent circuit and Eq. 9. The approximate value of $r_c/(1 - A)$ is thus $1/(1 - 0.94) = 16.6$ K.

If a resistor is connected between grid and cathode, its value must be multiplied by this same factor $1/(1 - A)$ in a calculation of the equivalent input resistance of the circuit. Sometimes a resistor is connected between grid and a tap on the cathode resistor.[1] Its apparent value is then increased in the same way, but by a smaller amount. The factor $1/(1 - A)$ may still be applied if A is now the ratio of changes of voltage at the tap point to changes of input voltage.

B. *R-C* CIRCUITS

Tubes and resistors alone cannot produce all of the waveforms required by radar sets. To determine the slope of a saw-tooth wave, the duration of a gate pulse, or the repetition frequency of a trigger pulse, some means of controlling the variation of a voltage with *time* is needed. A simple means is the charging or discharging of a capacitor through a resistor or a tube. Thus *R-C* circuits occur frequently in radar systems.

[1] The reason for using such a connection is explained in Art. 11.

4. Transients in *R-C* Circuits. Figure 8 illustrates the waveforms obtained in *R-C* circuits. Capacitor C in the circuit of Fig. 8*A* is assumed to be initially uncharged. Closing switch S causes the voltage e_i to jump suddenly from zero to the battery voltage E, as shown in Fig. 8*B*. Because the voltage across a capacitor cannot change instantly,[1] the entire voltage E appears across R at the first instant after S is closed, and none of it appears across C—see the waveforms of e_R and e_C. The voltage

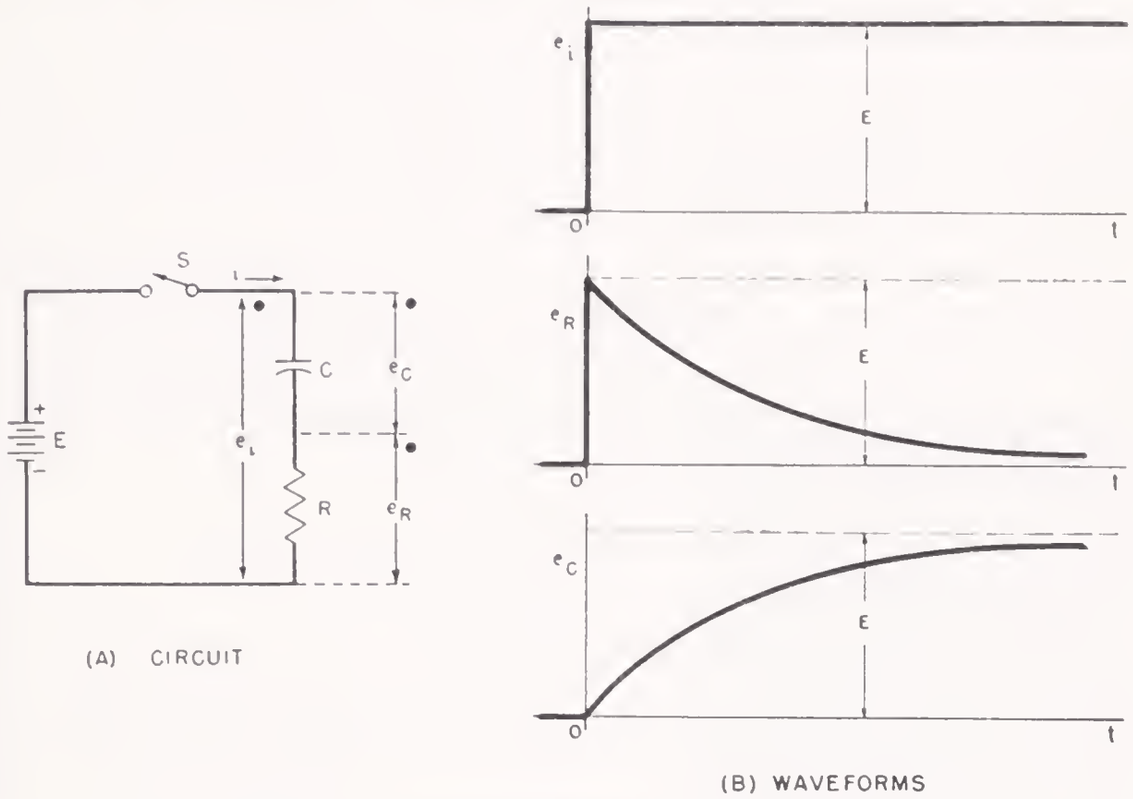


FIG. 8. Circuit illustrating *R-C* transients.

across R causes current to flow, and therefore, after the initial instant, C charges and e_C rises. As the charging continues, the voltage available to send current through the resistor decreases, and therefore the current and the rates of change of voltages e_R and e_C decrease. In fact, the slope of the e_R curve is at each point proportional to the ordinate because

$$\frac{de_R}{dt} = -\frac{de_C}{dt} = -\frac{1}{C}\frac{dq}{dt} = -\frac{i}{C} = -\frac{e_R/R}{C} = -\frac{1}{RC}e_R \qquad (11)$$

where q is the charge on C, $q = Ce_C$, and i is the current through the circuit.

Because e_R decreases with time in this particular manner, it may be represented by the exponential relation

$$e = E\epsilon^{-t/\tau} \qquad (12)$$

[1] Mathematically a zero-resistance battery suddenly connected to an uncharged capacitor would charge the capacitor instantly. However, all actual circuits have resistance; this resistance makes it impossible to change a capacitor voltage instantly.

where

e is the value of an exponentially varying voltage at any time t,

E is the value of e at the initial instant $t = 0$,

ϵ is the base of natural logarithms, 2.718 . . . ,

τ is a constant, called the *time constant.*

Differentiation shows that the graph of $e = E\epsilon^{-t/\tau}$, like the curve of e_R, has a slope proportional to its ordinate, for

$$\frac{de}{dt} = -\frac{E}{\tau}\epsilon^{-t/\tau} = -\frac{1}{\tau}e \tag{13}$$

Therefore e_R is an exponential voltage. Comparing Eqs. 11 and 13 shows that the time constant of e_R is

$$\tau = RC \tag{14}$$

A sketch of $e = E\epsilon^{-t/\tau}$ is drawn as an insert in Fig. 9. Exponential transients occur so frequently that it is useful to memorize the values

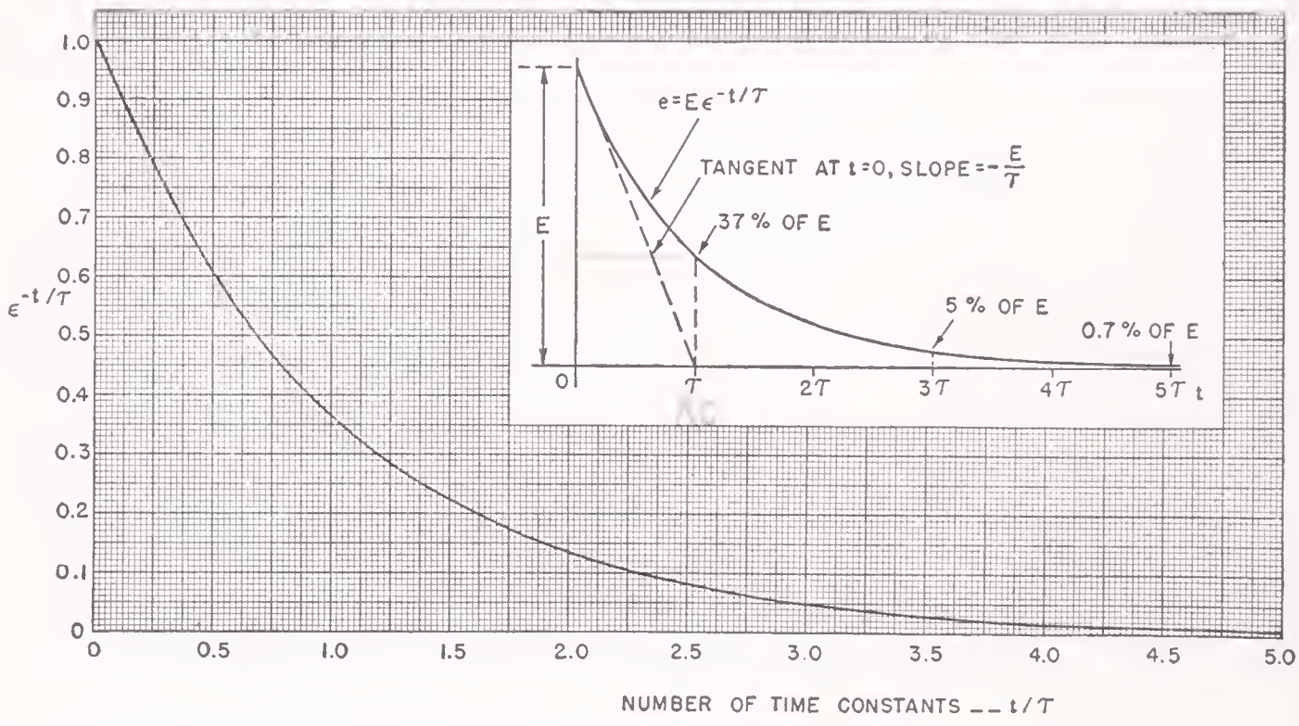


FIG. 9. The exponential function $\epsilon^{-t/\tau}$.

indicated on this sketch. A plot of Eq. 12 for unit value of E is included as the main part of Fig. 9. This plot is sufficiently accurate for use in rough calculation.

Equation 12 may be used to represent the capacitor voltage e_C as well as the resistor voltage e_R, provided e is considered to be the difference between e_C and its final value E. In fact this one equation (or the one curve of Fig. 9) may be used to determine points on *any* exponential transient, provided *e and E are always considered as deviations from the final value.*[1]

[1] Numerical examples of this method of calculating points on exponential curves are given in later articles, for example, Arts. 5, 9, and 13.

Any R-C circuit containing only *one* capacitor and any number of resistors, batteries, and constant-current sources may be changed by Thévenin's theorem to a circuit like that of Fig. 8*A*. Therefore, *all* voltage and current transients in such circuits are exponential in form and have a time constant RC, where C is the capacitor value and R is the equivalent resistance of Thévenin's theorem; that is, R is the net resistance connected to the terminals of C (measured with all voltage sources short-circuited and current sources open-circuited). Similarly, if a circuit contains one resistor and any number of capacitors, only exponential transients can occur. If, however, two or more of each kind of element are present, transients in the circuit may be the sum of two or more exponentials of different time constants. The voltage and current variations then depend on the relative time constants and amplitudes of the component exponentials.

The following facts about R-C circuits are so important that they should be memorized:

1. The current-voltage relation for a capacitor is

$$i = C\frac{de}{dt} \tag{15}$$

2. The voltage across a capacitor cannot change instantly.
3. The average current through a capacitor is zero.[1]
4. Transients in single-capacitor R-C circuits are always exponential with time constant RC.

5. Saw-tooth Generator. A simple and important circuit is the saw-tooth generator of Fig. 10*A*. It forms the basis of nearly all range-sweep circuits and of many range-measuring circuits as well. The input voltage may have the form of e_c in Fig. 10*B* or may be any wave that maintains the tube cutoff for one time interval and conducting with $e_c = 0$ for a second time interval.

Before the time t_1, the tube is conducting and the plate voltage is low. At time t_1 the tube is cut off, and capacitor C charges through the load resistor R_L. Thus e_b increases exponentially toward 300 volts. At time t_2 the tube again conducts and quickly discharges the capacitor, reducing e_b to its original low value. Thus the saw-tooth waveform of e_b in Fig. 10*B* is obtained. The rise of voltage is nearly linear, provided the time interval between t_1 and t_2 is short enough to allow e_b to rise to only a small

[1] A precise statement is that the average current is zero if the capacitor voltage is the same at the beginning and end of the averaging period. Average current is of interest in circuits where the capacitor voltage changes periodically. In these circuits the average capacitor current during any interval including many cycles is very nearly zero and the average over an integral number of cycles is exactly zero.

fraction of the supply voltage, so that only the very beginning of the exponential is used.

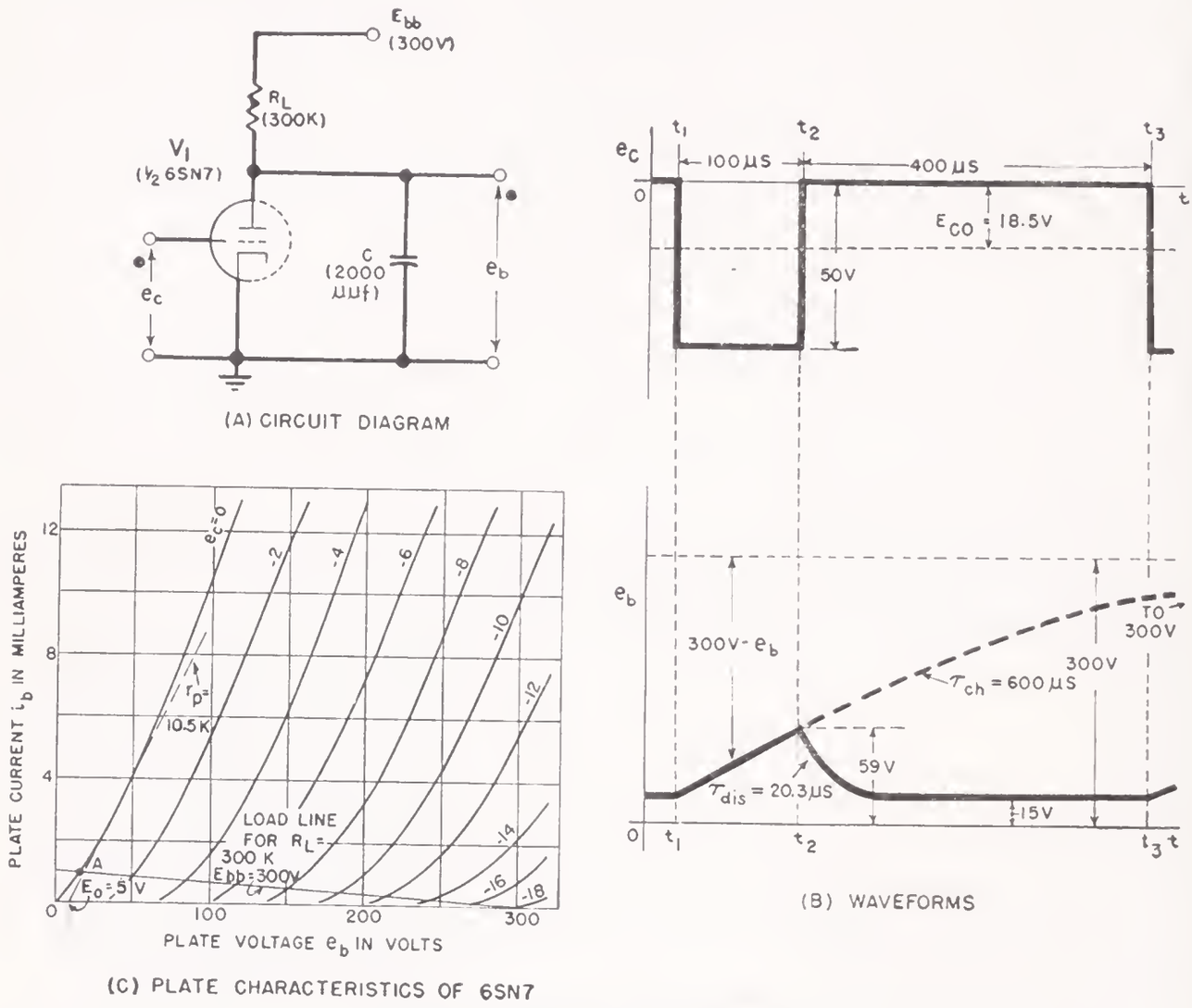


FIG. 10. Saw-tooth generator.

Typical values for the circuit constants are indicated in parentheses in Fig. 10*A*. Assume that immediately before time t_1 transients resulting from previous operations have died out and C has discharged to the final value. This assumption will be justified later. The plate current of the tube must then equal the current in R_L. Hence, a 300-K load line may be drawn, as in Fig. 10*C*, and the plate current and plate voltage determined from point A. At this point $i_b = 0.95$ ma and $e_b = 15$ volts.

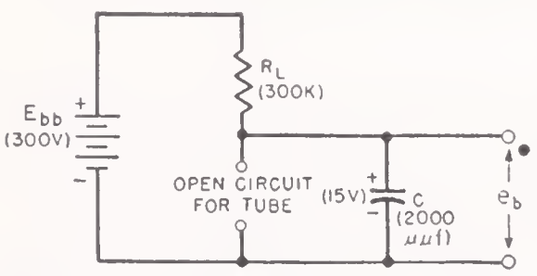


FIG. 11. Equivalent circuit for Fig. 10*A* during cutoff period.

At time t_1 the tube is cut off,[1] and the circuit of Fig. 10*A* becomes equivalent to the simpler circuit of Fig. 11. The voltage (15 volts) indicated in magnitude and polarity beside C is the *initial* voltage across

[1] The actual grid voltage is −50 volts; −18.5 volts would produce cutoff even if $e_b = 300$ volts. Actually e_b does not rise so high, and a smaller magnitude voltage would ensure cutoff.

C for the cutoff period; it is equal to the capacitor voltage just before time t_1, because the voltage across C cannot change instantly.

From Fig. 11, it may be seen that the voltage e_b rises exponentially from 15 toward 300 volts, as indicated in Fig. 10*B*. The time constant of the exponential is[1]

$$\tau_{ch} = (300 \times 10^{-3}) \times 2{,}000 = 600 \ \mu s$$

where the subscript *ch* is applied because 600 μs is the time constant of the charge of capacitor C. The value of e_b at time t_2 may be found by considering $300 - e_b$ as an exponentially varying voltage:

$$300 - e_b = (300 - 15)\epsilon^{-100/600} = 241 \text{ volts}$$
$$e_b = 300 - 241 = 59 \text{ volts}$$

How nearly linear the rise from 15 to 59 volts is may be judged by calculating the final voltage that would be reached if the rise were perfectly linear. Suppose the rise continued at the initial rate, which is 285 volts in one time constant (600 μs); the voltage e_b after 100 μs would be $15 + 285 \times {}^{100}\!/_{600} = 62.5$ volts. Thus the actual curve falls away 3.5 volts from the linear curve.

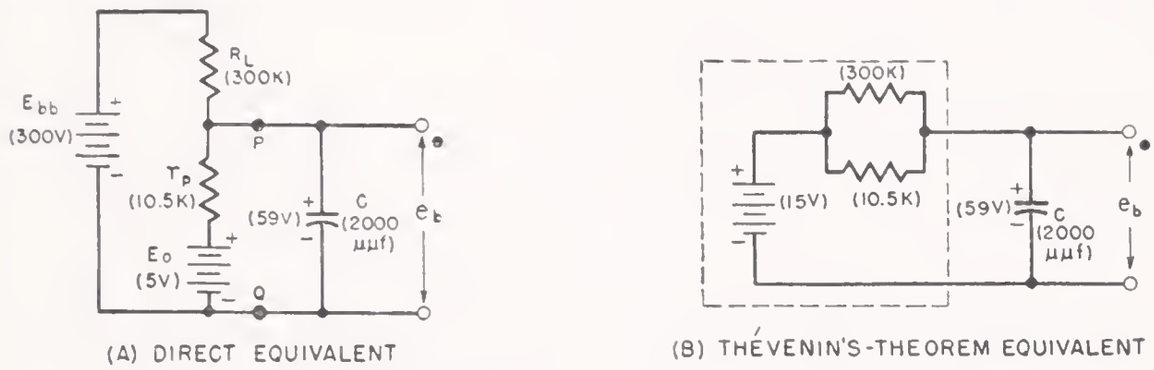


FIG. 12. Equivalent circuits for Fig. 10*A* during conducting period.

At time t_2 the grid voltage returns to zero and the tube conducts. The equivalent circuit becomes that of Fig. 12*A*. In this figure, 59 volts is the initial capacitor voltage equal to the capacitor voltage at the end of the cutoff period. Since $e_c = 0$ throughout the conducting period, the resistance and battery representing the plate-cathode path are determined from the dashed line in Fig. 10*C* (see Fig. 4*A* and the discussion on page 60). This line is drawn to coincide as nearly as possible with the part of the $e_c = 0$ characteristic of the tube used by this circuit. This part is between $e_b = 15$ volts and $e_b = 59$ volts, and the line drawn yields the values $E_o = 5$ volts and $r_p = 10.5$ K.

To simplify calculations, Thévenin's theorem may be used to replace the part of the circuit to the left of points P and Q in Fig. 12*A* by the

[1] In calculating short time constants, convenient units to use are *megohms* and *micromicrofarads*. The time constant is obtained in microseconds.

circuit within the dashed rectangle of Fig. 12*B*. The battery voltage in Fig. 12*B* is the open-circuit voltage between points *P* and *Q* which is

$$5 + \frac{(300 - 5) \times 10.5}{10.5 + 300} = 15 \text{ volts}$$

As a matter of fact, the determination of E_o and the calculation from it of the open-circuit voltage are entirely unnecessary. The open-circuit voltage between points *P* and *Q* is the plate voltage of the conducting tube when the capacitor current is zero and was found from point *A* in Fig. 10*C* to be 15 volts. The equivalent circuit of Fig. 12*B* is therefore completely determined by the *R*-*C* elements in the actual circuit, the value of e_b at point *A*, and the slope of the dashed line in Fig. 10*C*.

After time t_2, e_b decreases exponentially from 59 to 15 volts. The time constant of the exponential obtained from Fig. 12*B* is

$$\tau_{dis} = \left(\frac{300 \times 10.5}{300 + 10.5} \times 10^{-3}\right) \times 2{,}000 = 20.3 \ \mu\text{s}$$

Because τ_{dis} is many times smaller than the time interval available for discharge (20.3 μs $\ll$ 400 μs), the discharge is for all practical purposes complete before the next cycle begins, and the original assumption (e_b = 15 volts at time t_1) is justified.

Immediately after time t_2, the plate current is considerably greater than the final value of 0.95 ma. The initial plate current may be determined from the initial plate voltage (59 volts) and the $e_c = 0$ line of the tube characteristics in Fig. 10*C*. The current is 5.25 ma. This relatively large value results because at first the tube carries the capacitor discharge current, in addition to the small current, from the power supply through the load resistor. As the capacitor discharges, the tube current decreases and eventually equals the load-resistor current.

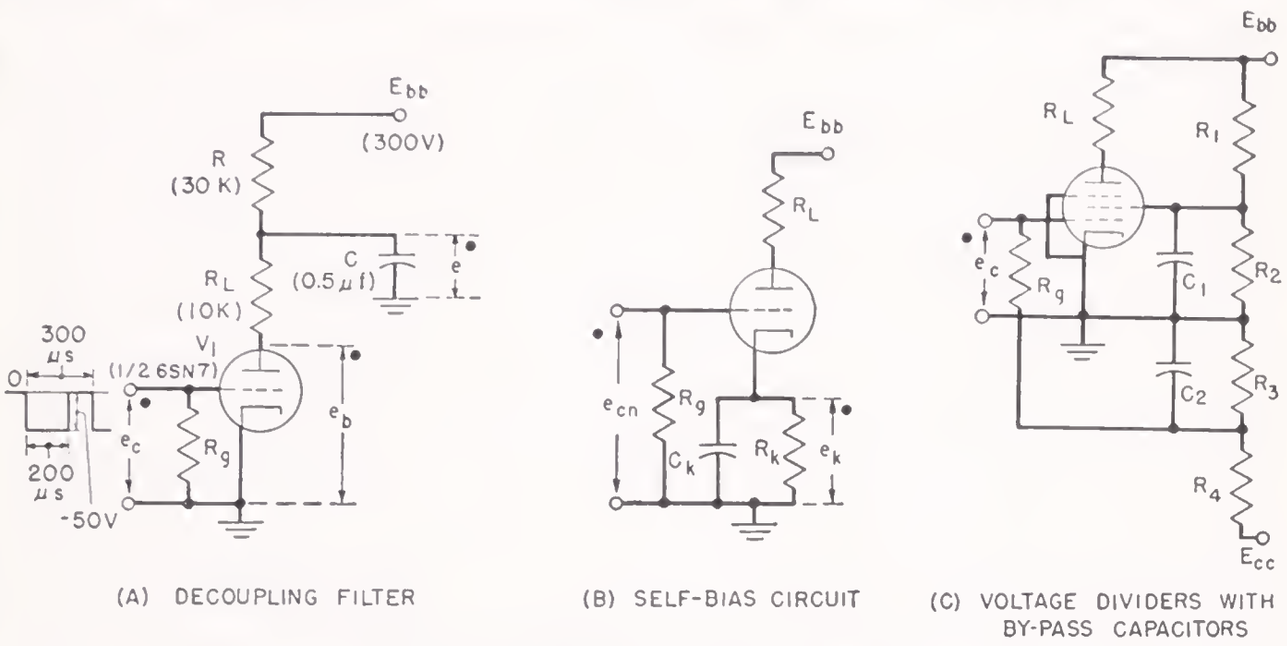
A caution in the use of Thévenin's theorem should be observed. In Fig. 12*B*, the battery and resistors within the dotted rectangle are *not* physical elements. The initial current in the 10.5-K resistor, for example, can be calculated and is 4.2 ma. However, 4.2 ma is *not* the plate current or any other current in the actual circuit. Figure 12*B* can be used correctly only for calculation of currents and voltages in the circuit *outside* the dashed rectangle.

Note that the time constant for discharging *C* is much shorter than the charging time constant. The difference results because the high resistance in the charging circuit (300 K) is shunted by a relatively low resistance (10.5 K) in the discharging circuit. A long time constant for charge is necessary in order to obtain a nearly linear saw tooth. A short time constant for discharge is required if a second saw-tooth wave is to be started very soon after the first ends and if each saw tooth is to begin with *C* discharged to its final voltage. The time interval between t_2 and t_3 is

not used; if e_b is a range-sweep voltage, the unused interval is the period between sweeps.

6. Decoupling Filters and Bias Circuits. Proper operation of tubes requires that a variety of grid-bias, screen-supply, and plate supply voltages be available in a radar set. Resistance-capacitance networks of the type illustrated in Fig. 13 are commonly used to derive the needed voltages from a few main-power-supply voltages.

All the capacitors in Fig. 13 are large, so that the voltages across them cannot change greatly during a cycle of operation of the circuit. Thus, as a first approximation, each capacitor may be replaced by a battery, as is



(A) DECOUPLING FILTER (B) SELF-BIAS CIRCUIT (C) VOLTAGE DIVIDERS WITH BY-PASS CAPACITORS

FIG. 13. Typical R-C circuits for altering power-supply voltages.

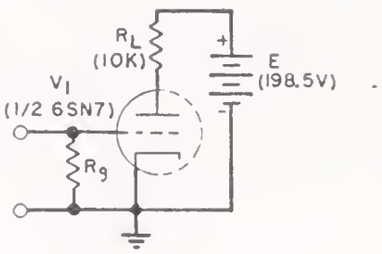
done in Fig. 14. The resistors and average tube currents in Fig. 13 determine the voltages of these batteries, and the capacitors determine the extent to which the "battery" voltages are constant throughout the cycle.

Resistance-capacitance networks like those of Figs. 13A and C serve also to reduce the reaction of one circuit on another when both are operated from the same power supply. For example, if the tube in Fig. 13A draws plate current in short pulses, the pulses are supplied by the capacitor which recharges slowly between pulses. A nearly constant current is drawn from the power supply, which may therefore deliver a constant voltage to other circuits.

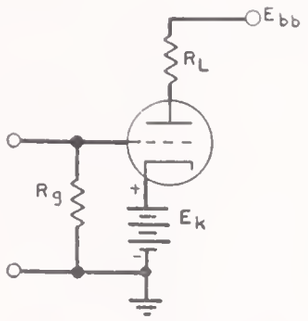
As an example of the details of the operation of such circuits, consider Fig. 13A and assume that the circuit elements and the input wave e_c are as indicated on the figure. The capacitor voltage e may be found by a trial-and-error calculation if it is assumed to be essentially constant because of the large value of C.[1]

[1] Frequently by-pass capacitors are not large enough to make this assumption valid. Qualitative explanation may still be made in terms of the equivalent circuits of Fig. 14, but calculation of exact values requires use of the method applied to coupling circuits in Art. 9.

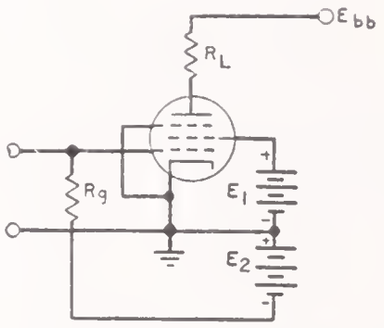
Assume values of e_b during the conducting interval and calculate E_{bb}, trying new assumptions until the given value, 300 volts, is obtained for E_{bb}. The final trial is $e_b = 97$ volts, and the final calculation is as follows: Since $e_c = 0$, the plate current obtained from the tube curves is $i_b = 10.15$ ma. Since the average current through the capacitor must be zero, the *constant* current in R must equal the *average* tube current.



(A) DECOUPLING FILTER



(B) SELF-BIAS CIRCUIT



(C) VOLTAGE DIVIDERS WITH BY-PASS CAPACITORS

FIG. 14. Equivalent circuits for Fig. 13 if capacitors are large enough to maintain their terminal voltages constant.

Since the tube conducts for 100 μs out of 300 μs, this average current is $10.15 \times {}^{100}\!/_{300} = 3.38$ ma. Therefore,

$$\begin{aligned} E_{bb} &= e_b + i_b R_L + i_{b_{avg}} R \\ &= 97 + 10.15 \times 10 + 3.38 \times 30 = 300 \text{ volts} \end{aligned}$$

To justify the assumption that the capacitor voltage is constant, the charge gained by the capacitor during the cutoff interval may be calculated. The capacitor current during this interval equals the current in R, 3.38 ma. The charge gained is $(3.38 \times 10^{-3}) \times 200 = 0.676$ μcoulomb, and the voltage change is $0.676/0.5 = 1.35$ volts. The capacitor voltage, therefore, varies approximately 1 volt about an average value of $97 + 10.15 \times 10 = 198.5$ volts.

Trial-and-error calculations may be made equally well for the circuits of Figs. 13*B* and *C*. In the self-bias circuit, the bias voltage is R_k times the average cathode current. In the circuit of Fig. 13*C*, the voltage across C_1 may be found by applying Thévenin's theorem to the voltage divider (R_1 and R_2) and considering the average screen-grid current to be drawn from it.

If the input voltage for an overdriven circuit is changed, the voltages delivered by circuits like those in Fig. 13 may change greatly. For example, if the repetition period of the 200-μs negative pulse, e_c in Fig. 13*A*, were changed from 300 to 210 μs, the voltage e would rise from 198.5 to 279 volts. Such effects must be considered in radar sets where the repetition frequency or the range scale may be changed.

7. *R-C* Peaker. To produce narrow pulses from rectangular waves or from other waveforms having sudden jumps in them, an *R-C* peaker may

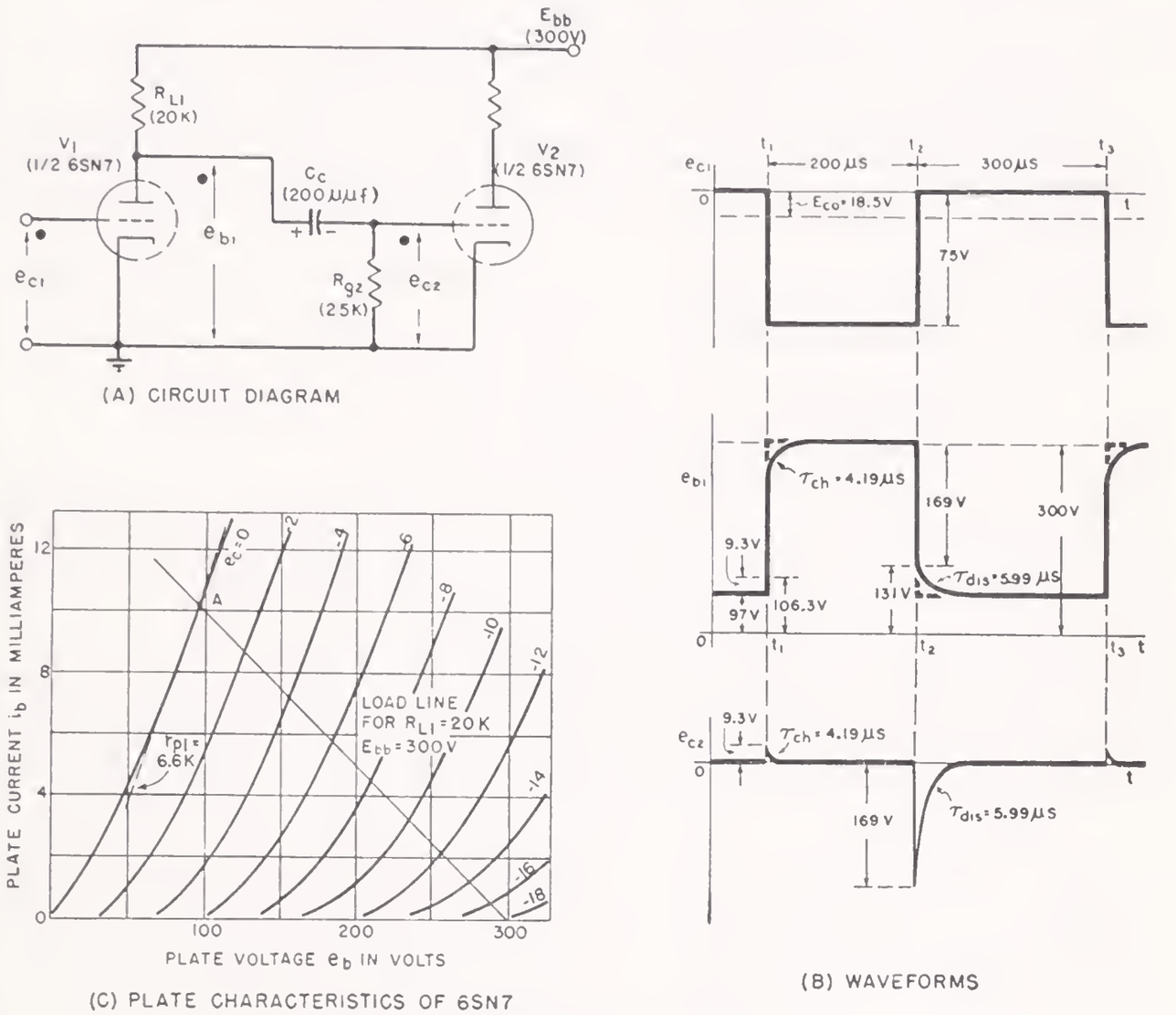


FIG. 15. Example of *R-C* peaker circuit.

be used. The pulses produced are suitable for triggering gas tubes, multivibrators (see Arts. 13 to 15), and blocking oscillators (see Art. 21) or may be clipped to form narrow pulses of nearly rectangular shape. Peaker circuits are sometimes called *sharpening circuits* or *differentiating circuits*.

An example of an *R-C* peaker circuit is shown in Fig. 15*A*. Note that the circuit contains the same elements as the ordinary coupling circuit; the difference lies in the size of capacitor C_c. For a peaker circuit, this capacitor (and often the grid-return resistor R_{g2} as well) is made small so

that the time constants of transients are small fractions of the off and on periods of tube V_1.

To study the circuit in detail, assume the values of circuit elements indicated in parentheses in Fig. 15*A* and the input wave e_{c1} drawn in Fig. 15*B*. The load line for R_L may be drawn and used to find i_{b1} and e_{b1} at the end of the conducting period of V_1, because the short time constants of the circuits ensure that the capacitor current is zero at this time. Thus $i_{b1} = 10.15$ ma, and $e_{b1} = 97$ volts. Also, because the capacitor current is zero, e_{c2} must be zero, and the voltage across C_c must be 97 volts with the polarity indicated in Fig. 15*A*.

At time t_1 tube V_1 is suddenly cut off, and the equivalent circuit of Fig. 16 applies during the cutoff interval. The static grid resistance of V_2 is included in Fig. 16 because cutting off V_1 tends to make e_{b1} and therefore e_{c2} rise. The initial voltage for the cutoff period is indicated below the capacitor.

Ohm's law may be applied to the circuit of Fig. 16 to determine the current i at the initial instant:

$$\text{Initial current} = \frac{300 - 97}{20 + (25 \times 1)/(25 + 1)} = 9.68 \text{ ma}$$

Therefore, when tube V_1 is cut off the current in the tube stops instantly, but the current in the load resistor merely drops from 10.15 to 9.68 ma. This current flows through capacitor C_c and the grid circuit of V_2 and, of course, dies out very quickly as C_c charges.

From the initial current, the initial values of e_{c2} and e_{b1} may be calculated. The initial grid voltage is $9.68 \times (1 \times 25)/(1 + 25)$, or 9.3 volts, and the initial plate voltage is the sum of e_{c2} and the capacitor voltage, or $9.3 + 97 = 106.3$ volts. These values are marked in Fig. 15*B*. Note that the 9.3-volt rise of e_{b1} at the instant of cutoff must equal the corresponding rise of e_{c2} because the voltage across C_c cannot change instantly.

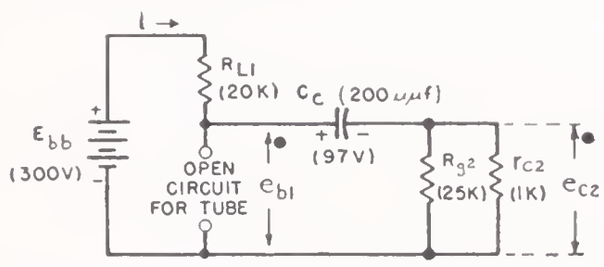


FIG. 16. Equivalent circuit for Fig. 15*A* during cutoff period of V_1.

Following the initial instant, e_{b1} rises exponentially toward 300 volts, and e_{c2} falls exponentially toward zero as C_c charges. The time constant of these exponentials, which is the charging time constant for C_c, is obtainable from Fig. 16:

$$\tau_{ch} = \left(20 + \frac{1 \times 25}{1 + 25}\right) \times 10^{-3} \times 200 = 4.19 \ \mu\text{s}$$

By the end of the cutoff period for V_1, the voltage across capacitor C_c has reached its final value of 300 volts. Therefore in the equivalent

circuit for the conducting period of V_1 (see Fig. 17*A*), C_c appears with an initial voltage of 300 volts. The plate-to-cathode path of V_1 is represented by r_{p1} and E_{o1}; and the value of r_{p1}, determined from the dashed line in Fig. 15*C*, is 6.6 K. The dashed line is made to approximate the $e_c = 0$ curve in a short range above point *A* because this is the range used when V_1 conducts. The resistor r_{c2} is omitted because e_{c2} is negative during the conducting period of V_1.

A Thévenin's theorem equivalent of Fig. 17*A* appears in Fig. 17*B*. Note that the 97-volt battery is the open-circuit voltage between ter-

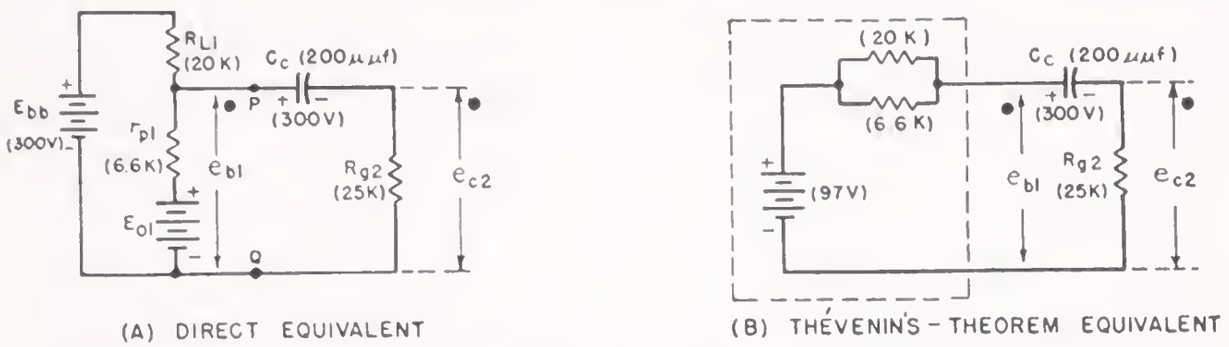


FIG. 17. Equivalent circuits for Fig. 15*A* during conducting period of V_1.

minals P and Q in the direct equivalent circuit. By Ohm's law, the initial current in the Thévenin's equivalent circuit is

$$\text{Initial current} = \frac{300 - 97}{25 + (20 \times 6.6)/(20 + 6.6)} = 6.78 \text{ ma}$$

This current is in the direction to discharge C_c. The initial value of e_{c2} is therefore negative and is $-6.78 \times 25 = -169$ volts. The corresponding plate voltage e_{b1} is $-169 + 300 = 131$ volts. The initial plate current, obtained from the plate voltage and the tube characteristic, is 15.1 ma, which is greater than the steady-state current, 10.15 ma. As in the saw-tooth generator, the plate current is the result of the battery voltage and the capacitor voltage and decreases as the capacitor discharges.

The time constant for the capacitor discharge and for the exponential return of e_{b1} to 97 volts and e_{c2} to zero is

$$\tau_{dis} = \left(25 + \frac{20 \times 6.6}{20 + 6.6}\right) \times 10^{-3} \times 200 = 5.99 \ \mu\text{s}$$

Two observations should be made about the waveforms in Fig. 15*B*. The first is that the negative peak in the e_{c2} wave is much larger than the positive peak. This effect is mostly due to the current drawn by the grid of V_2 when its voltage is positive and may be thought of as a form of grid-circuit clipping. It is partly due to the driving amplifier which has greater internal resistance when V_1 is cutoff than when V_1 conducts. The second observation is that the waveform of the plate voltage of V_1 is

not the rectangular dotted wave in Fig. 15*B*, but a wave with rounded corners. The tube may be regarded as a generator with internal resistance—its terminal voltage changes when a load (the peaker circuit) is applied.

8. Sloping Wavefront Applied to Peaker. In Art. 7, the tube that drives the peaker is assumed to be turned on and off by a perfectly rectangular input wave. Waves with perfectly vertical sides can never be obtained in actual circuits; therefore, the effect of sloping sides on the peaker output should be investigated.

Consider, first, the circuit of Fig. 18*A* with the linearly rising voltage e_i of Fig. 18*B* applied to it. Capacitor C is supposed to be uncharged initially. The output voltage e_o is an exponential of time constant

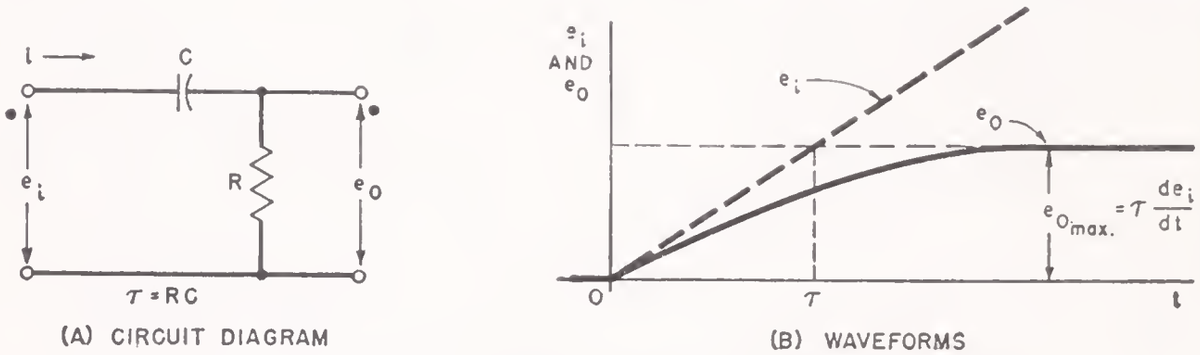


FIG. 18. Linearly increasing voltage applied to *R-C* peaker.

$\tau = RC$ rising to the final value $\tau de_i/dt$, as shown in Fig. 18*B*. Note that (1) the final value of e_o is equal to the change of the input voltage in one time constant and (2) the final value of e_o is proportional to the rate of change of the input voltage. Statement (2) is the justification for the name *differentiating circuit* sometimes applied to the peaker.

The reason that the output voltage approaches a constant value may be seen as follows: Immediately after $t = 0$, e_o must rise with e_i, for the voltage across C cannot change instantly. As time progresses, e_o increases but becomes increasingly less than the input voltage because the capacitor charges. In fact, as e_o becomes greater, the current i and therefore the rate of change of the capacitor voltage become greater and less of the rate of change of the input voltage becomes available for e_o. Hence the slope of the e_o curve becomes progressively less. When e_o reaches the value $\tau de_i/dt$, the current through the resistor is

$$\frac{e_o}{R} = \frac{\tau de_i/dt}{R} = C\frac{de_i}{dt}$$

The capacitor current $C \times de_{\text{capacitor}}/dt$ must equal the resistor current Cde_i/dt, and therefore the rate of change of capacitor voltage equals the rate of change of input voltage, and e_o ceases to change. The increase of e_o from zero to $\tau de_i/dt$ is exponential in form because it is a transient in a single-capacitor *R-C* circuit.

The general effect of sloping sides for the e_i wave in Fig. 15*B* can be seen if, for simplicity, the grid resistor R_{g2} in Fig. 15*A* is assumed to be much larger than R_{L1}. Then the peaker has negligible effect upon the plate voltage e_{b1}, and the load line can be used to determine the plate voltage when V_1 begins to conduct. Suppose that a wave like e_{b1} in Fig. 19*A* is produced. The time interval T corresponds to the time for e_{c1} to rise from the cutoff voltage to zero. During this interval, conditions are like those in Fig. 18; at the end of it, e_{b2} becomes constant, and an exponential decay of e_{c2} occurs as indicated in Fig. 19*A*. The shape and amplitude of e_{c2} depend upon the relation of T to the peaker time constant τ, as indicated in the three parts of Fig. 19.

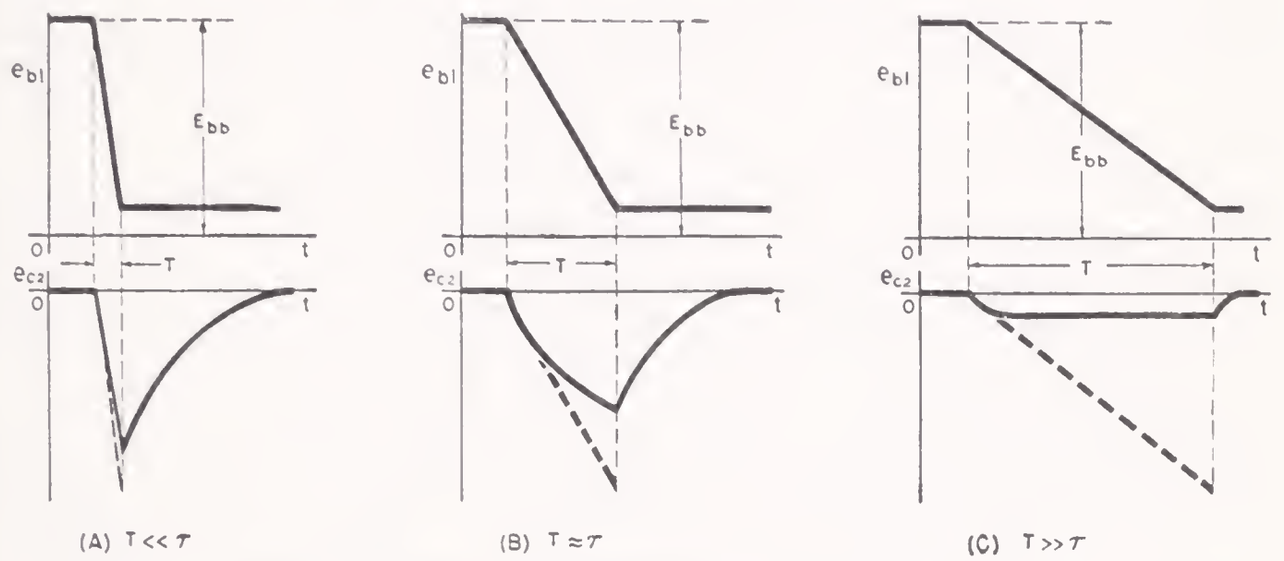


FIG. 19. Effect of *R-C* peaker on a sloping wavefront.

Sometimes a peaker is used to convert a saw-tooth voltage into a nearly rectangular pulse. The time constant of the peaker must be short compared with the duration of the saw tooth. The waveforms obtained are then essentially those of Fig. 19*C*.

9. *R-C* Coupling Circuit. To couple a rectangular pulse or other timing waveform from the circuit in which it is generated into the circuit in which it is used, long-time-constant *R-C* circuits are commonly employed. As with coupling circuits in sine-wave amplifiers, the object ordinarily is to change the d-c component of the wave without modifying the alternating components.

The circuit of Fig. 20*A*, for which the input voltage wave e_{c1} is drawn in Fig. 20*B*, is a typical coupling circuit. Note that this circuit and input wave are the same as the peaker circuit and input wave in Fig. 15, except that in the coupling circuit C_c and R_{g2} are increased in value. The waveforms in the coupling circuit therefore resemble the peaker waveforms but have all the exponential variations spread out in time, as in Fig. 20*B*. Because of the long time constants, the plate voltage never rises all the way to 300 volts while the tube is cut off; the coupling capacitor is still

charging when the next switching operation occurs. Similarly, during the conducting period the capacitor does not have time to complete its discharge, and the plate voltage does not drop all the way to 97 volts. Also, the grid voltage fails to reach zero at the end of each conducting period.

Since final values are not attained within each part of the cycle, the exact values of currents and voltages during each period depend upon the voltage to which the capacitor was charged at the end of the previous

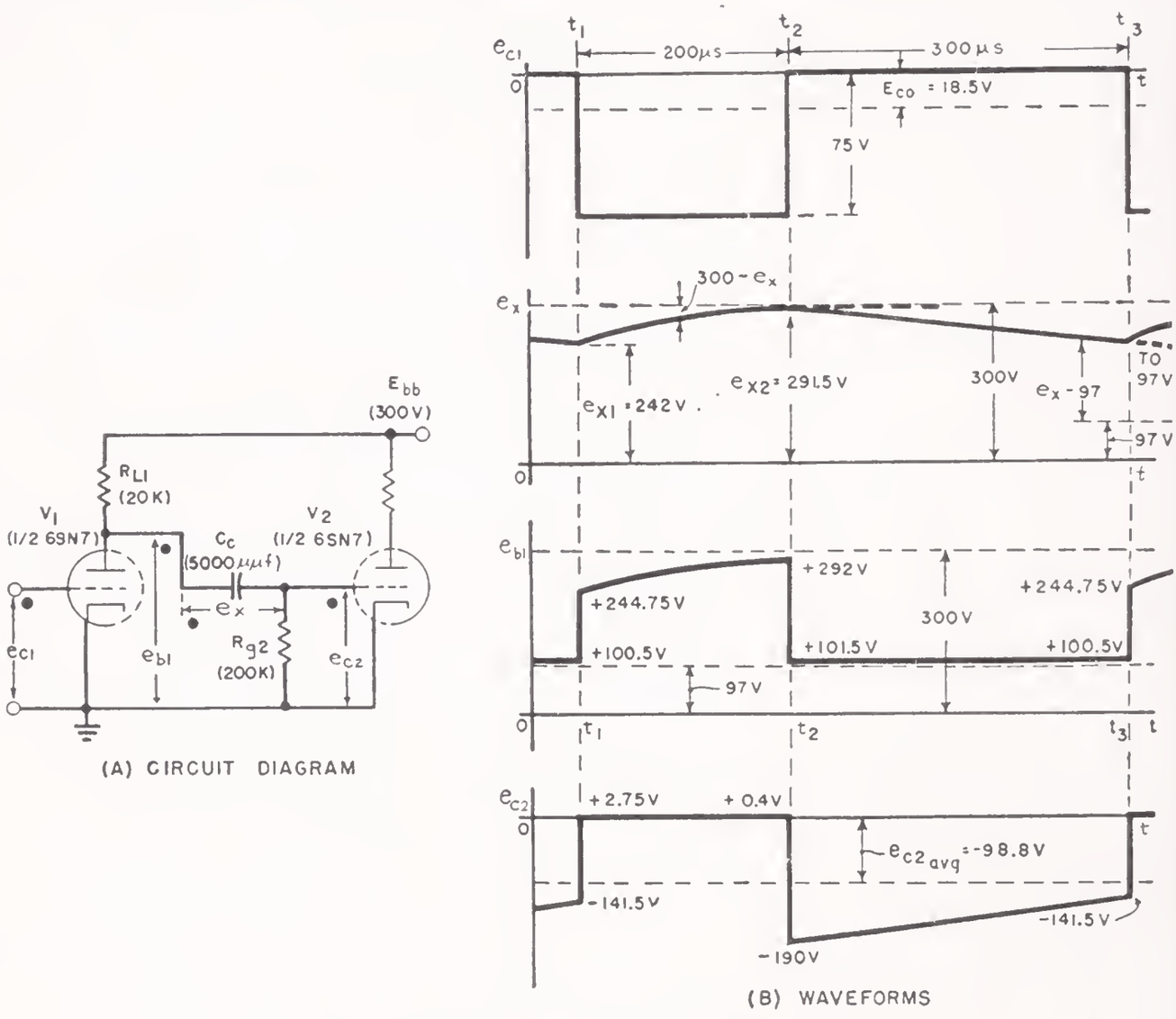


FIG. 20. Typical *R-C* coupling circuit.

period. When the circuit is first turned on, a few cycles of operation occur during which the capacitor gains a little more charge during each charging period than it loses in the ensuing discharging period. Thus the average capacitor voltage increases, and each cycle is a little different from the preceding one. After a number of cycles, a repeating condition is reached, and the capacitor gains and loses equal amounts of charge. The waveforms of Fig. 20*B* apply to this condition.

In order to calculate points on the e_{b1} and e_{c2} waveforms, values of the capacitor voltage e_x at the switching instants must first be found. Designate the value of e_x at time t_1 by e_{x1} and the value at time t_2 by e_{x2}. The

equivalent circuit for the cutoff period of V_1, which is identical with Fig. 16 except for values of R_{g2} and C_c, may be used to calculate e_{x2} in terms of e_{x1}. Since the final value toward which e_x is rising during the cutoff period is 300 volts, the exponential relation is

$$300 - e_{x2} = (300 - e_{x1})\epsilon^{-200/105} = 0.149(300 - e_{x1}) \tag{16}$$

as may be seen from Fig. 20*B*. The number 200 is the time interval between t_1 and t_2, and 105 is the time constant, both in microseconds. In a similar way, the equivalent circuit for the conducting period of V_1 (see Fig. 17*B*) may be used to calculate the value of e_x at time t_3 (which voltage is equal to e_{x1} because the repeating condition is assumed to have been reached) in terms of e_{x2}. The final value toward which e_x is dropping is 97 volts, and the time constant obtained from the equivalent circuit is 1,025 μs. Therefore

$$e_{x1} - 97 = (e_{x2} - 97)\epsilon^{-300/1,025} = 0.746(e_{x2} - 97) \tag{17}$$

Solving Eqs. 16 and 17 simultaneously yields the values 242 volts for e_{x1} and 291.5 volts for e_{x2}. These values, together with the time constants, completely determine the waveform of e_x.

Calculation of e_{b1} and e_{c2}, once e_x is known, is the same as the corresponding calculation for the peaker circuit. For example, the initial value of e_{c2} after V_1 is cut off is found by applying Ohm's law to the equivalent circuit for the cutoff period. Thus

$$\text{Initial } e_{c2} = (300 - 242)\frac{(200 \times 1)/(200 + 1)}{20 + (200 \times 1)/(200 + 1)} = 2.75 \text{ volts}$$

During the cutoff period for tube V_1 the increase of the coupling-capacitor voltage causes e_{b1} to rise much more than e_{c2} falls because the plate load resistor (20 K) is large compared with the grid resistance (approximately 1 K). The grid voltage is essentially constant at zero; it does not become appreciably positive even momentarily (as in the peaker) because the initial capacitor voltage is nearly as large as the supply voltage. During the conducting period of V_1, the plate-circuit resistances are low relative to the grid-circuit resistances; therefore e_{b1} is nearly constant and almost equal to 97 volts, while e_{c2} rises as the capacitor discharges.

Note and remember the following important points:

1. The larger the time constants of the circuit compared with the periods to which they apply, the more nearly are the plate and grid waveforms alike—except for d-c components. (Increasing the size of C_c, for example, would reduce the rise of e_{c2} between times t_2 and t_3. In this circuit, however, there is no reason for such a change because as long as e_{c2} remains below cutoff, the output of V_2 is unaffected.)

2. The area between the zero axis and the positive part of the e_{c2} wave is very small compared with the area between the axis and the negative part of the wave. That is, the *average value of e_{c2} is negative*, even though no source of direct current is present in the grid circuit. This negative average value is called *signal bias* or *grid-leak bias;* it is discussed in detail in the next article.

10. Signal Bias and Clampers. Signal bias occurs in the circuit of Fig. 20*A* because the charging resistance for C_c between the grid of V_2 and ground is different from the corresponding discharging resistance. During the time that e_{c2} is positive, capacitor C_c charges through the low resistance of r_{c2} in parallel with R_{g2}. During the remainder of the cycle, the capacitor must lose all the charge acquired in this first part; to do so, it must discharge through the high resistance of R_{g2}. Therefore, the negative grid voltage occurring during discharge must be of greater magnitude than the positive voltage applied during charge (or the negative voltage must be applied for a longer time).

The amount of the signal bias may be calculated in terms of the change of charge on capacitor C_c. Between times t_1 and t_2, the voltage across C_c in Fig. 20*A* increases by 291.5 − 242 = 49.5 volts, and the charge on the capacitor increases by an amount $49.5 \times (5{,}000 \times 10^{-12}) = 0.248$ μcoulomb, which is the area under the curve of current through the capacitor during the time interval t_1 to t_2. Since this current flows through r_{c2} and R_{g2} in parallel to produce e_{c2}, the area under the positive part of the e_{c2} curve must be $0.248 \times [10^3 \times (200 \times 1)/(200 + 1)] = 247$ volt-μs. During the time interval between t_2 and t_3, this same charge flows in the opposite direction through R_{g2}. Thus the area under the negative part of the e_{c2} wave is $-0.248 \times (200 \times 10^3)$ volt-μs, which equals −49,600 volt-μs. The average value of e_{c2} is therefore

$$e_{c2_{avg}} = \frac{\text{area under 1 cycle of curve}}{\text{period of 1 cycle}}$$

$$= \frac{247 - 49{,}600}{500} = -98.8 \text{ volts}$$

If the ratio of R_{g2} to r_{c2} is very large, as in Fig. 20*A*, the area under the positive part of the e_{c2} wave is negligible compared with that under the negative part. Then it is an excellent approximation to assume that the *topmost part of the e_{c2} wave lies along the zero axis.* In such circumstances, the top line of the wave is said to be *clamped* to zero.

Often diodes are used as clampers in place of the grid-cathode circuit of V_2 for the specific purpose of clamping a waveform. Thus in Fig. 21*A*, the diode V_2 is used to clamp the output of the saw-tooth-generator tube V_1. Because the *cathode* of the diode is connected to the signal lead, the *bottommost* line of the e_o waveform is clamped to −50 volts, the voltage

applied to the plate. If any part of the e_o waveform should become negative with respect to −50 volts, the diode would conduct and over a period of several cycles charge C_c and restore the waveform to the position shown. Note that the position of the bottommost edge of the waveform is essentially independent of the remainder of the wave. For example, if the duration of the saw tooth is changed (by changing the duration of the input pulse and switching from capacitor C_1 to C_2), the clamping point is not appreciably affected, as shown by the dotted lines in Fig. 21*A*. Because clamper tubes restore d-c components to waveforms that have passed through a coupling capacitor, clampers are often called *d-c restorers*.

In the circuit of Fig. 21*B*, the triode V_2 is the clamper. This tube is equivalent to a diode with its plate returned to some positive voltage.

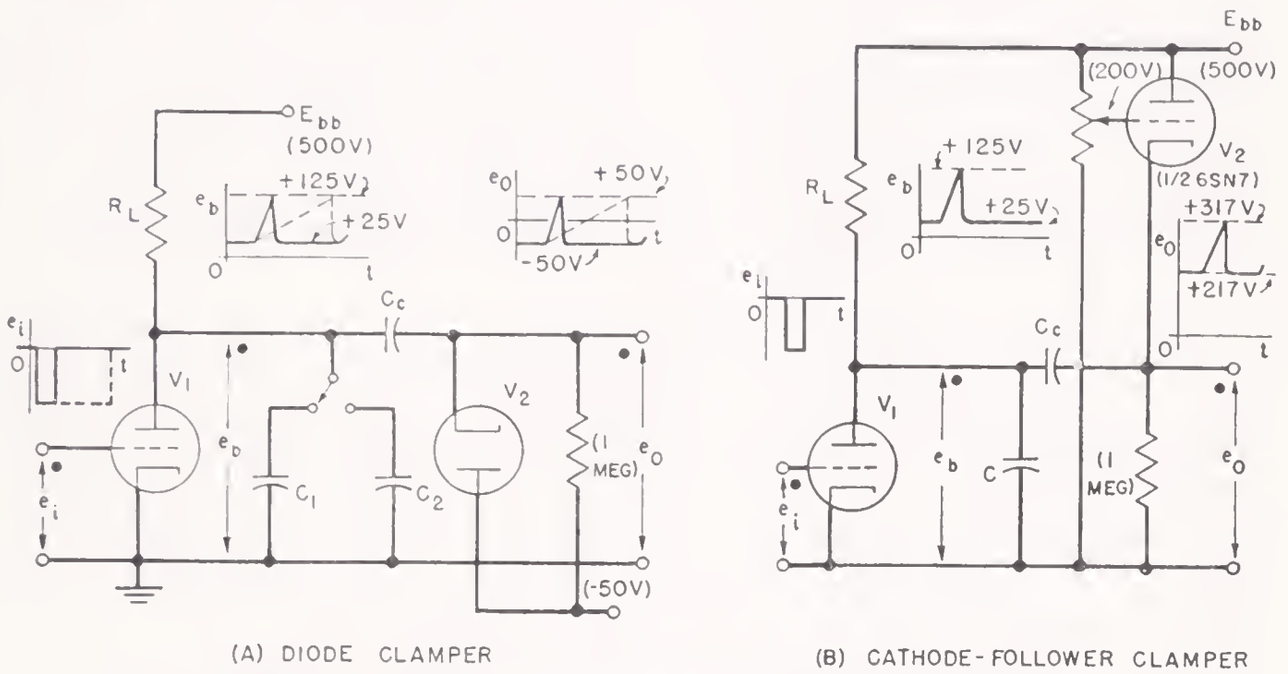


FIG. 21. Circuits containing clampers or d-c restorers.

If the cathode voltage is above a certain value, the tube conducts no current because the grid-to-cathode voltage is negative and beyond cutoff; cathode voltages below this value allow current to flow. The critical value of cathode voltage for cutoff may be found by trial and error. For the grid-to-ground voltage of 200 volts indicated in the figure, the result is 217 volts, because then the plate voltage for V_2 is 500 − 217 = 283 volts, and the corresponding grid-to-cathode cutoff voltage (read from the 6SN7 characteristics) is −17 volts. Therefore the waveform of the output voltage e_o is as shown. Any tendency of e_o to become less than 217 volts causes V_2 to conduct and to change the charge on capacitor C_c. The voltage level at which clamping occurs can be changed readily by adjusting the tap on potentiometer R.

11. Effect of Grid Return. In the coupling circuit of Fig. 20*A*, the grid-return resistor R_{g2} is connected from grid to cathode of V_2. Some-

times this resistor is returned to a different point, and the waveforms obtained are somewhat altered. The method of calculation given in Art. 9 may still be used, but care must be taken to use equivalent circuits appropriate to the connection under consideration.

If the grid return is to a point of sufficiently negative voltage, as in Fig. 22*A*, the grid voltage for tube V_2 may be entirely below the zero axis as shown. Then grid current does not flow at any point in the cycle, the average current in R_{g2} must be zero, and the average value of the e_{c2} wave is the bias voltage (-100 volts in the example). Between times

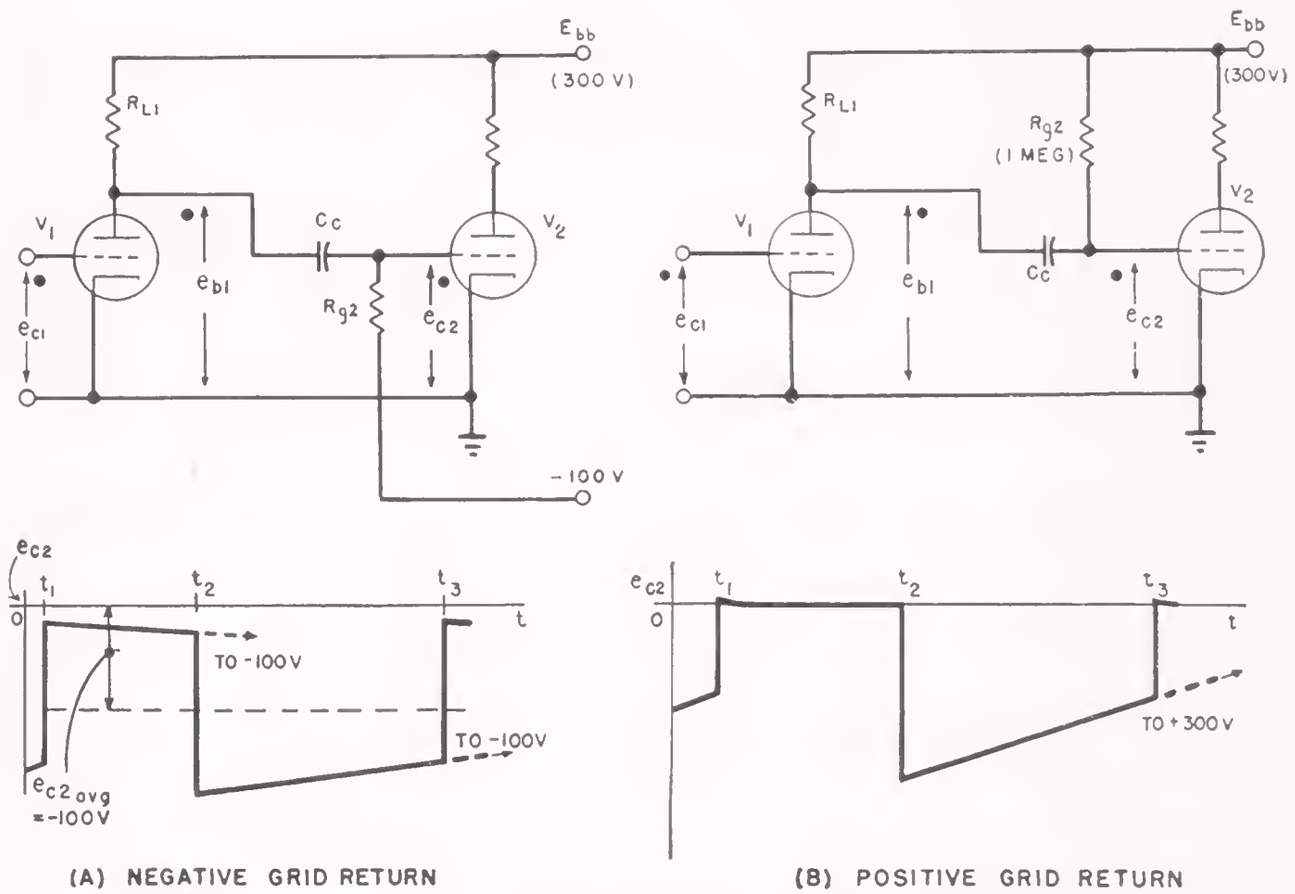


FIG. 22. Effect of grid-return connection.

t_1 and t_2, e_{c2} falls exponentially toward a final value of -100 volts as C_c charges, because the current in and voltage across R_{g2} are decreasing toward zero during the discharge. Similarly, the exponential curve that forms the bottom edge of the waveform tends toward a final value of -100 volts. The absence of grid current allows appreciable variation of e_{c2} between time t_1 and t_2 and correspondingly less variation of e_{b1}. (Compare the waveforms of e_{c2} in Figs. 20*B* and 22*A*.)

Frequently the grid return is made to a large positive voltage, for example, the plate supply voltage as in Fig. 22*B*. In such a circuit the grid-return resistor must always have a very large value (often 0.5 to 5 megohms) in order to prevent damaging the tube by excessive grid current.

Consider first that the positive-grid-return connection is made but no signal voltage is being supplied by V_1. Then a current flows through the grid-return resistor and the grid resistance in series. By the voltage-divider relation for $R_{g2} = 1$ megohm and $r_{c2} = 1$ K, the grid voltage is $300 \times 1/(1{,}000 + 1) = 0.3$ volt. Thus in static operation, the positive grid return makes the grid voltage only slightly positive.

When a signal is supplied by V_1, e_{c2} still is driven only very slightly positive *during the most positive part of the cycle* because by signal-bias action additional charge is accumulated on C_c. During the interval between t_1 and t_2, the charging current for C_c and the current in R_{g2} together make up the grid current for V_2 and cause e_{c2} to become positive by a few tenths of a volt. During the remainder of the cycle, grid current does not flow, but the current in R_{g2} is increased. This current is the discharging current for C_c and causes the exponential rise of e_{c2}. Note that e_{c2} rises toward a final value of +300 volts, the voltage of the return point for R_{g2}. If the return connection of R_{g2} is moved from cathode to a point of positive voltage without changing the value of any circuit element, the *time constant of the rise of e_{c2} is not changed but the rate at which e_{c2} rises is increased* because the exponential is rising toward a higher final voltage.

Figure 23 shows examples of grid-return connections that are used with cathode followers. The return connection to the cathode (Fig. 23*A*) corresponds to the amplifier circuit of Fig. 20*A*; the topmost part of the grid-voltage wave is clamped at a grid-to-cathode voltage of zero. As shown in Art. 3, resistors R_g and r_c are equivalent to larger resistors from grid to a point of fixed voltage, and therefore the time constants of exponentials are increased. The positive-grid-return connection for the cathode follower appears in Fig. 23*B*. This circuit operates in the same manner as that of Fig. 22*B*. If the amplitude of the input voltage e_i is not too great, the circuit of Fig. 23*C* is like the negative grid return of Fig. 22*A* because the positive voltage at the cathode prevents the grid from drawing current. Sometimes a bias voltage intermediate to those obtained from Figs. 23*A* and *C* is desired. Then a tap on the cathode resistor R_k may be used as in Fig. 23*D*. Until the grid-current point is reached, moving the tap point upward on R_k merely increases the d-c component of the voltage e_1 by changing the voltage across C_c. If the tap point is moved close to the cathode, or if the variations of e_i are large, grid current may flow during part of the cycle and cause clamping to occur.

If clipping in a cathode follower is to be avoided, the circuit of Fig. 23*C* should be used when the input voltage is a series of narrow positive pulses, and the circuit of Fig. 23*D* should be employed when the input is a series of narrow negative pulses.

12. Step Generator. Voltage waveforms containing vertical edges or *steps* adjustable in position may be used as explained in Art. 6, Chap. I, in the measurement of range. Clipping circuits operating in conjunction with saw-tooth generators are often used to produce such waveforms. The step waveform is combined with echo signals and applied to the vertical-deflection plates of the indicator tube (see p. 239, Chap. IV).

A simple step-generating circuit, illustrated in Fig. 24*A*, employs a saw-tooth generator, tube V_1, and a *pick-off diode*, tube V_2. Tube V_1

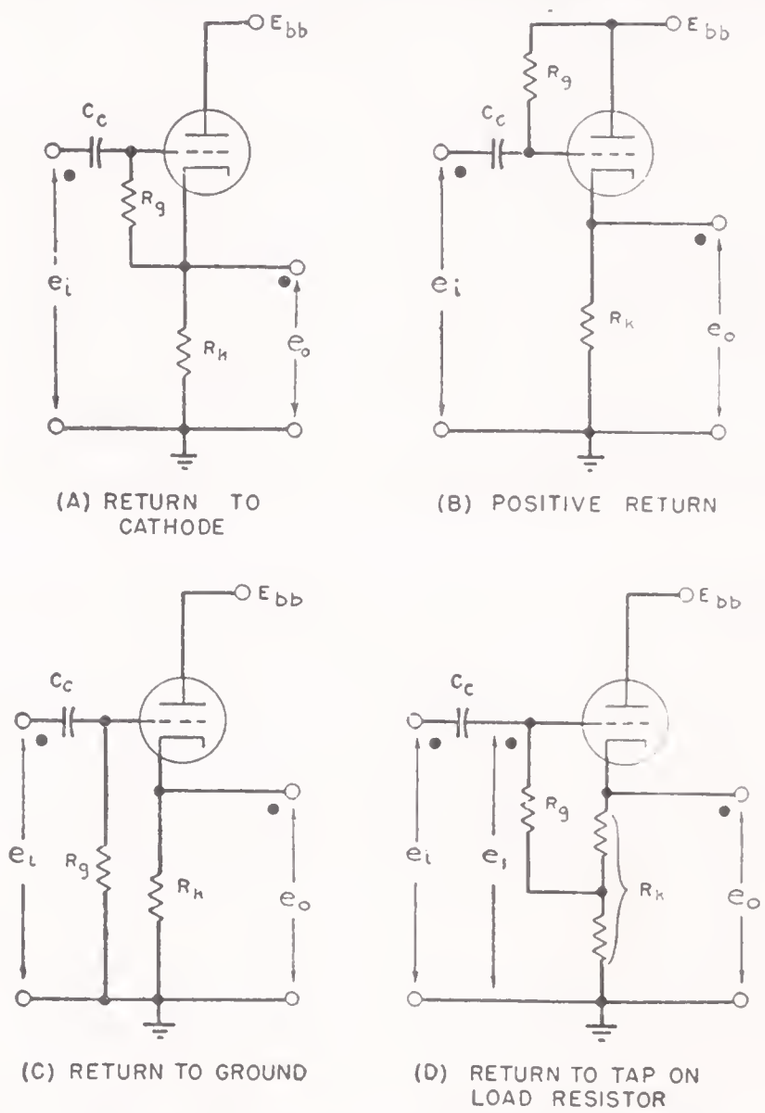


FIG. 23. Grid-return connections used with cathode followers.

produces the e_{b1} waveform of Fig. 24*B* when the rectangular input wave e_{c1} is applied. The horizontal dashed line across the e_{b1} waveform indicates the voltage at the adjustable tap on resistor R. When e_{b1} is less than this voltage, the diode does not conduct and e_d equals the voltage at the tap. For higher values of e_{b1}, the diode conducts and e_d very nearly equals e_{b1}. Thus the top corner of the saw tooth appears in the e_d wave as in Fig. 24*B*. The e_d wave is clipped and amplified so that a nearly vertical edge appears in the e_o waveform at the instant V_2 begins to conduct. The time of the step can therefore be controlled by adjustment of the tap on R. A linear relation exists between Δt, the delay of the step, and the

tap voltage if the saw tooth is linear. Calibrations on the dial that controls the tap position can be used to indicate the range that corresponds to the delay time of the step.

The circuit elements in the saw-tooth generator are assumed to be the same as those in the circuit of Fig. 10A, p. 70. Therefore, in Fig. 24 the e_{b1} waveform previous to the time of conduction of the diode is the same

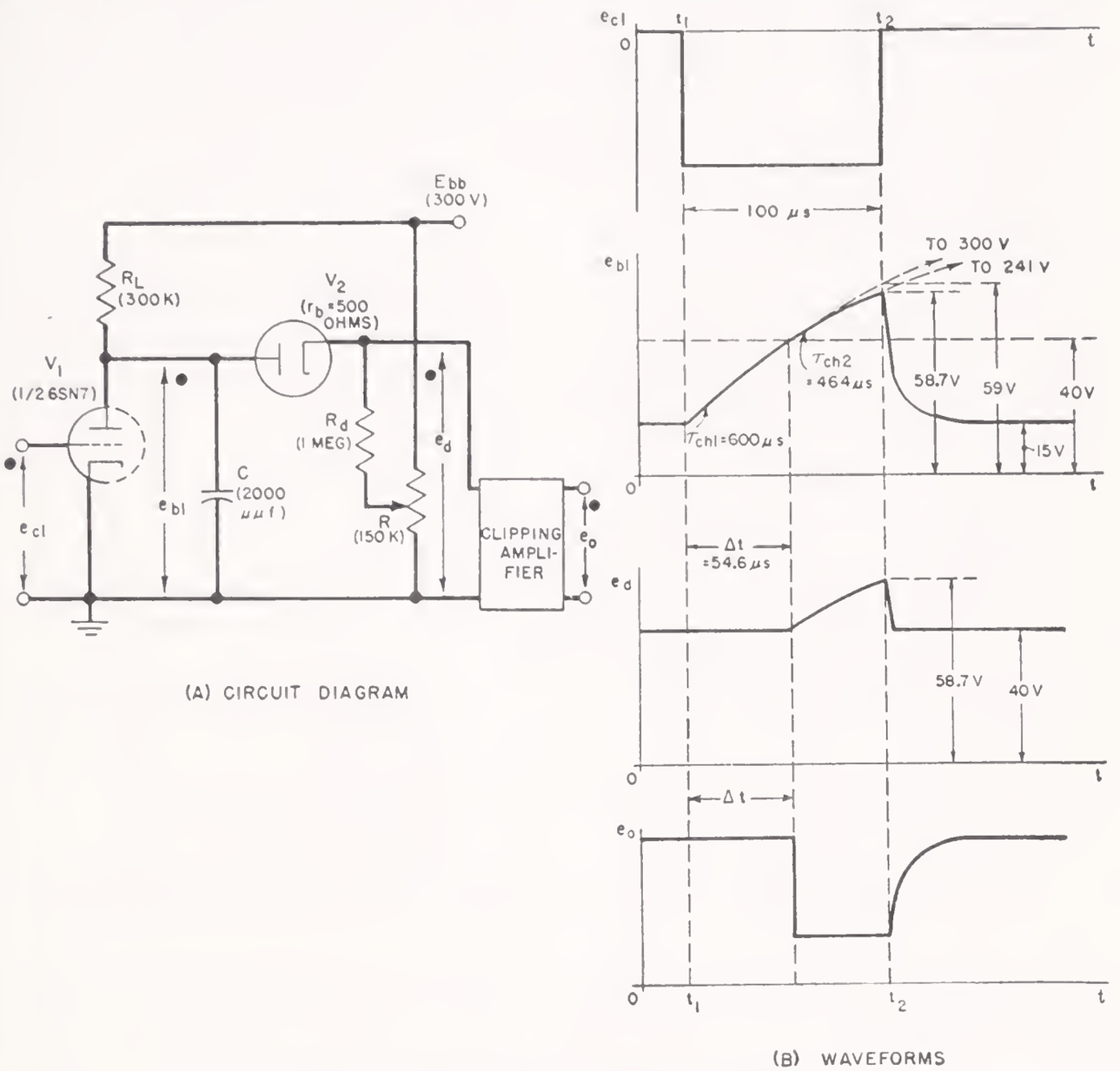


FIG. 24. Step generator using pick-off diode.

as the corresponding part of the e_b wave in Fig. 10*B*. Assume the resistance of the portion of R below the adjustable tap to be 20 K. The voltage represented by the dashed line on the e_{b1} waveform is then $300 \times {}^{20}/_{150} = 40$ volts, and the delay of the step may be obtained from the exponential relation

$$300 - 40 = (300 - 15)\epsilon^{-\Delta t/600} \tag{18}$$

and is $\Delta t = 54.6$ μs.

Exponential relations similar to Eq. 18 may be used to determine Δt for any setting of the adjustable tap. The over-all relationship between Δt

and the angle of rotation of the delay-control shaft depends upon (1) the relation of Δt to the open-circuit voltage at the tap, (2) the relation of this open-circuit voltage to the resistance in the lower section of R, and (3) the relation of this resistance to the angle of rotation of the shaft. The voltage-resistance relation is necessarily linear, and the resistance-angle relation can be made very closely linear if a properly designed potentiometer is used as R. The time-voltage relation depends upon the shape of the saw tooth and, for the circuit of Fig. 24*A*, is appreciably nonlinear (see Art. 5, p. 69). Where accurate range measurement and linear dial calibrations are desired, more complicated saw-tooth generators (such as those of Art. 6, Chap. IV) are employed to furnish saw-tooth waveforms of excellent linearity.

The form of the e_{b1} wave during the conducting time of the diode, and consequently the form of the e_d wave, is altered slightly because of current drawn by the diode. This current has no effect upon the time at which the step occurs, but it may be important if the e_{b1} wave serves as a sweep voltage in addition to being used in the step generator.

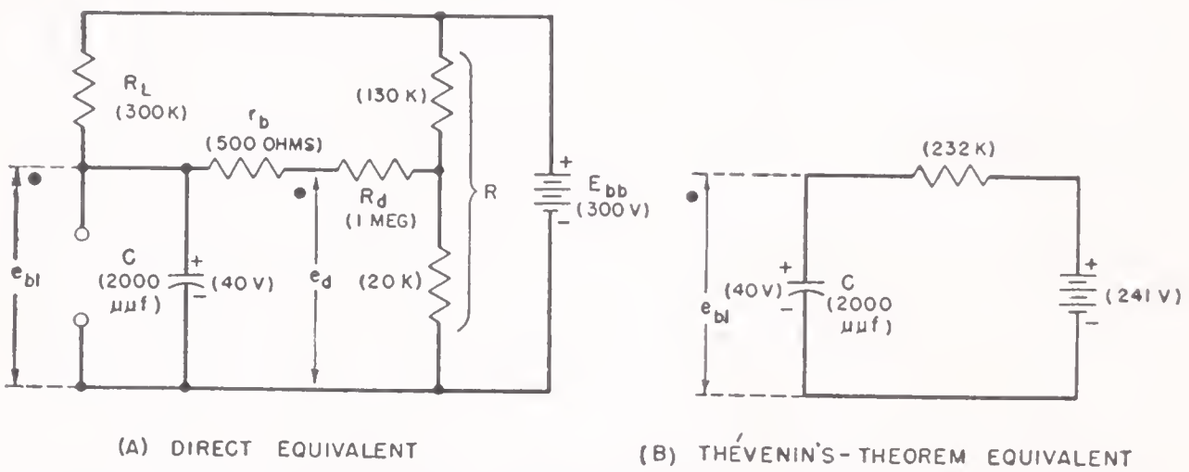


FIG. 25. Equivalent circuits for Fig. 24*A* with V_1 cutoff and V_2 conducting.

The equivalent circuit of Fig. 25*A* applies when the diode conducts. The assumption that the adjustable tap is set for 20 K in the lower section of R is still used, and therefore the initial capacitor voltage is 40 volts. Thévenin's theorem may be used to reduce the circuit of Fig. 25*A* to the one in Fig. 25*B*. The time constant of charge of capacitor C, according to Fig. 25*B*, is reduced from $\tau_{ch1} = 600$ μs to

$$\tau_{ch2} = 0.232 \times 2{,}000 = 464 \ \mu s$$

when the diode conducts. The final voltage toward which C is being charged is reduced from 300 to 241 volts. Therefore the value attained by e_{b1} at the end of the 100-μs period of the saw tooth is given by

$$241 - e_{b1} = (241 - 40)\epsilon^{-(100-54.6)/464}$$

and is 58.7 volts instead of 59 volts—a negligible difference.

The voltage e_d obtained from the diode when it is conducting is shown equal to e_{b1} in Fig. 24*B*—the small voltage across the diode is neglected. Calculations made by applying Thévenin's theorem to the portion of Fig. 25*A* to the right of the e_d terminals yield a diode voltage less than 0.01 volt when e_{b1} is 58.7 volts. Actually, a negative diode plate-to-cathode voltage of the order of 1 volt or less may be expected for small diode currents because of the initial velocities of electrons emitted from the cathode. The approximation of representing a diode by its average plate resistance is inaccurate in the region of very small currents and voltages.

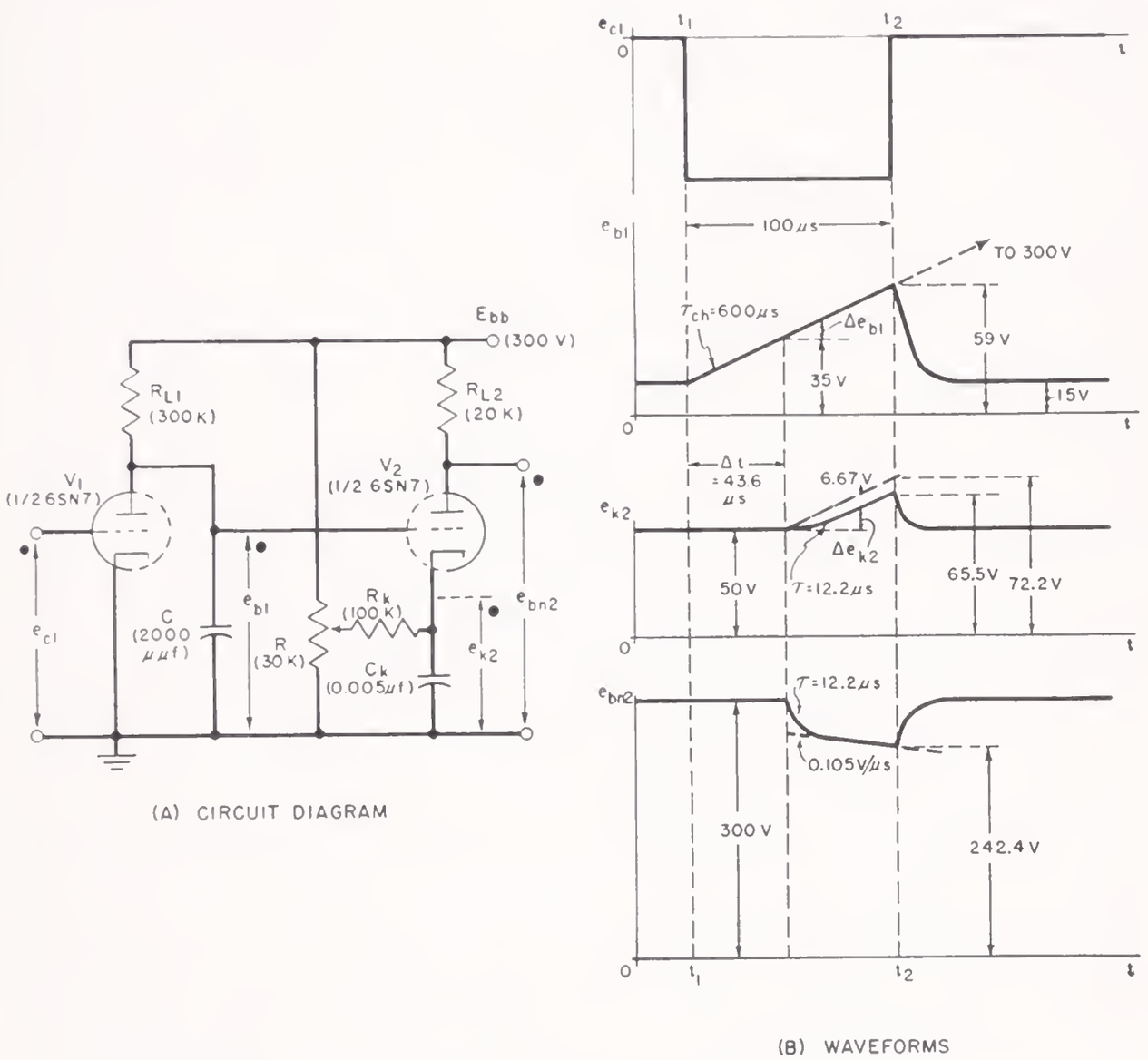


FIG. 26. Step generator using triode pick-off tube.

In some step generators a triode replaces the diode of Fig. 24*A*. The saw-tooth voltage is applied to the grid of the triode and the adjustable direct voltage to the cathode. The step is formed when the grid-cathode voltage rises above the cutoff value. One circuit using a triode pick-off tube is shown in Fig. 26*A*. This circuit is so designed that triode V_2 (1) serves as pick-off tube, (2) does not draw grid current and therefore

does not distort the saw-tooth voltage produced by V_1 and (3) contributes some of the amplification required to form the desired step.

Tube V_1 operates in the same circuit in Fig. 26*A* as in Figs. 10*A* and 24*A*. Because tube V_2 does not draw grid current, the saw-tooth voltage e_{b1} of Fig. 26*B* is identical with the e_b wave of Fig. 10*B*. Assume that the section of R below the adjustable tap is 5 K. When plate current in V_2 is cutoff (before and shortly after time t_1), the cathode voltage for V_2 is $e_{k2} = 300 \times \frac{5}{30} = 50$ volts, as indicated in Fig. 26*B*, and the plate-to-cathode voltage is 250 volts. Because the corresponding cutoff voltage (read from the 6SN7 curves) is 15 volts, plate current begins to flow in V_2 when e_{b1} rises to $50 - 15 = 35$ volts. Therefore Δt, determined from the exponential relation for the saw tooth, is 43.6 μs.

When plate current flows in V_2, the circuit operates as follows: At the time conduction begins, e_{k2} cannot change suddenly because of capacitor C_k. Therefore the rise of e_{b1} is at first fully effective in increasing the grid-cathode voltage, and the *plate-to-ground* voltage e_{bn2} decreases suddenly. As plate current flows into C_k, e_{k2} begins to rise along with e_{b1}. The rate of increase of the grid-cathode voltage is thus decreased, and further decrease of e_{bn2} occurs more gradually. Waveforms are drawn in Fig. 26*B*. Operation of this circuit is very similar to that obtained when a sloping wavefront is applied to an *R-C* peaker (see Art. 8). Voltage e_{k2} at first rises slowly because only a small plate current is available to charge C_k. As the plate current increases, e_{k2} increases more and more rapidly until the rate of increase of current in R_k becomes equal to the rate of increase of plate current. Thereafter the capacitor current is constant, and e_{k2} increases at a constant rate.

The purpose of C_k is to maintain e_{k2} nearly constant in the first instant of conduction so that the sharp decrease in the e_{bn2} wave is obtained, thus reducing the subsequent amplification and clipping required. The sharp drop of e_{bn2} could be obtained without C_k if R_k were short-circuited, but then e_{k2} would not rise at the end of the conduction period of V_2, and grid current would distort the e_{b1} wave.

Calculation of numerical values of voltages when V_2 conducts is simplified if the rise of voltage in the e_{b1} wave is assumed perfectly linear (with a slope of $\frac{59}{100} = 0.59$ volt per μs) and if the tube characteristics are assumed linear over the region used. If the resistance of the lower section of R is adjusted to 5 K as previously assumed, e_{b2} lies in a small range below 250 volts, i_{b2} in a range near 1 ma, and the approximations $\mu = 17$ and $r_p = 25$ K are reasonable for the 6SN7. The combination of V_2 and R_{L2} may be considered equivalent to a tube of plate resistance

$$r_p + R_{L2} = 25 + 20 = 45 \text{ K}$$

and amplification factor $\mu = 17$ connected as a cathode follower. The

equivalent circuit of Fig. 27*A*, obtained from Fig. 6*B*, p. 63, may therefore be used to relate Δe_{k2} (the increase of e_{k2} above 50 volts) to Δe_{b1} (the increase of e_{b1} above 35 volts).

By Thévenin's theorem, the circuit of Fig. 27*A* is equivalent to that of Fig. 27*B*. That is, the circuit is reduced to an *R-C* peaker with a sloping wavefront applied (see Art. 8), except that the output voltage e_{k2} is the voltage across the *capacitor* rather than that across the resistor. The waveform of e_{k2} is indicated by the solid-line curve in Fig. 26*B*. The time constant of the exponential, calculated from Fig. 27*B*, is

$$\tau = 2{,}440 \times 0.005 = 12.2\ \mu s$$

and the slope of the dashed line drawn on the e_{k2} waveform is 0.925 times the slope of the e_{b1} wave or $0.925 \times 0.59 = 0.546$ volt per μs. The limit-

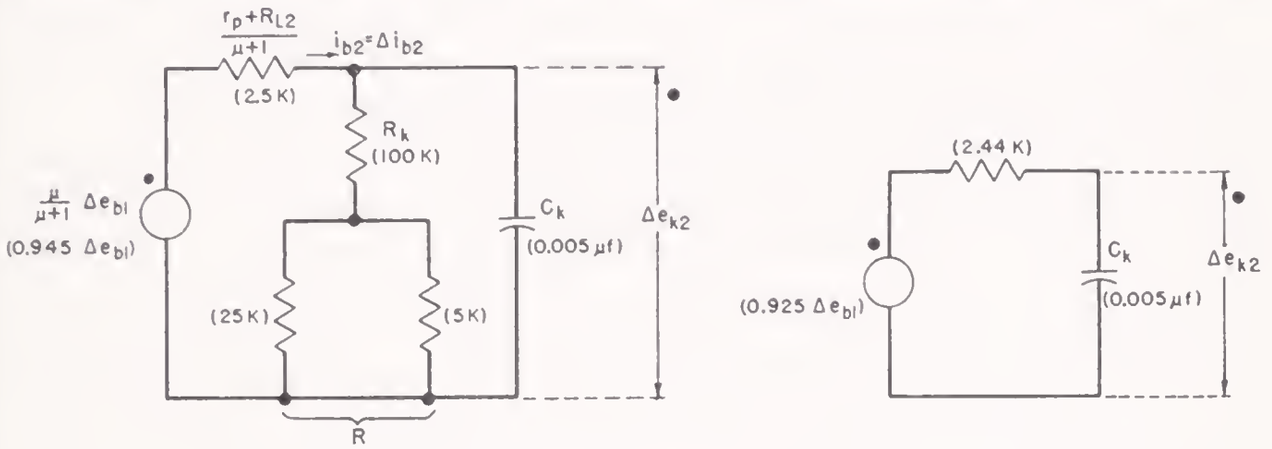


FIG. 27. Linear equivalent circuits for tube V_2 in Fig. 26*A*.

ing value of e_{k2} is less than the voltage of the dashed line because of the voltage drop in the resistor of Fig. 27*B*, which voltage drop is

$$\tau de_{k2}/dt = 12.2 \times 0.546 = 6.67 \text{ volts}$$

The voltages attained by the dashed line and by e_{k2} at time t_2 are therefore as indicated in Fig. 26*B*.

Note that e_{k2} is greater than e_{b1} at all points of the waveform. Thus the statement that grid current does not flow is justified for the assumed setting of R. A repetition of the above calculations for a tap position at the lower end of R shows that for the circuit parameters in Fig. 27*A* grid current is zero for this and, therefore, for all positions of the tap.

The waveform of e_{bn2} in Fig. 26*B* can be obtained by calculating i_{b2}, the plate current of V_2, and subtracting from E_{bb} the voltage across R_{L2} produced by this current. Shortly before time t_2 the current in capacitor C_k is constant, and the plate current is increasing at the same rate as current in R_k, that is, at

$$\frac{0.546}{100 + (25 \times 5)/(25 + 5)} = 0.00524 \text{ ma per } \mu s$$

The rate of decrease of e_{bn2} is therefore $0.00524 \times 20 = 0.105$ volt per μs. The plate current at time t_2 is the sum of the capacitor current

$$C \frac{de_{k2}}{dt} = 0.005 \times 0.546 = 0.00273 \text{ amp} = 2.73 \text{ ma}$$

and the resistor current

$$\frac{65.5 - 50}{100 + (25 \times 5)/(25 + 5)} = 0.147 \text{ ma}$$

This sum is 2.88 ma, and e_{bn2} at time t_2 is $300 - 20 \times 2.88 = 242.4$ volts. The final value and slope of e_{bn2} determine the position of the dashed line on the e_{bn2} waveform in Fig. 26*B*. The actual voltage e_{bn2}, shown by the solid line, is the sum of an exponential transient of time constant 12.2 μs and a linearly decreasing voltage; the actual voltage tends toward the dashed line as the transient tends toward its final value.

13. Free-running Plate-coupled Multivibrator. A circuit used in many places and in many forms in radar is the *multivibrator*. Most frequently, multivibrators produce nearly rectangular pulses of a specified (and sometimes adjustable) duration. The leading edge of each multivibrator pulse is timed by a driving trigger pulse. The trailing edge may be used to trigger another circuit, and thus a time delay may be obtained, or the multivibrator pulse may control a circuit like the saw-tooth generator of Art. 5. Less frequently, multivibrators are made to operate at a submultiple of the driving frequency or even allowed to run at a frequency determined by their own circuit constants.

Multivibrators differ from circuits previously discussed in that their cycle of operation does not depend instant by instant on a driving waveform. Some multivibrators go through a complete cycle of operation for each narrow trigger pulse received from a driving source and then remain quiescent until another trigger pulse is applied. These are called *one-cycle* multivibrators. In other multivibrators, plate current continually switches from one tube to another without the application of any driving voltage. These are *free-running* multivibrators.

The reason that multivibrators can determine their own cycle of operation is that feedback is used. The multivibrator circuit comprises two tubes with *R-C* coupling. If coupling is from the plate of each tube to the grid of the other tube, as in Fig. 28*A*, the multivibrator is said to be *plate coupled*. If one of the plate-to-grid coupling circuits is replaced by cathode-to-cathode coupling, as in Fig. 32*A*, p. 100, the circuit is a *cathode-coupled multivibrator*.

The free-running plate-coupled multivibrator of Fig. 28*A* is considered first. The figure shows the notation used: Subscripts 1 apply to tube

V_1 and its circuit, and subscripts 2 to tube V_2. Capacitor subscripts depend upon the grid to which the capacitor is connected.

When the power-supply voltages are applied to this multivibrator, currents begin to flow in the plate circuits of the tubes. Also charges build up on the capacitors as the plate voltages of the tubes increase. If the

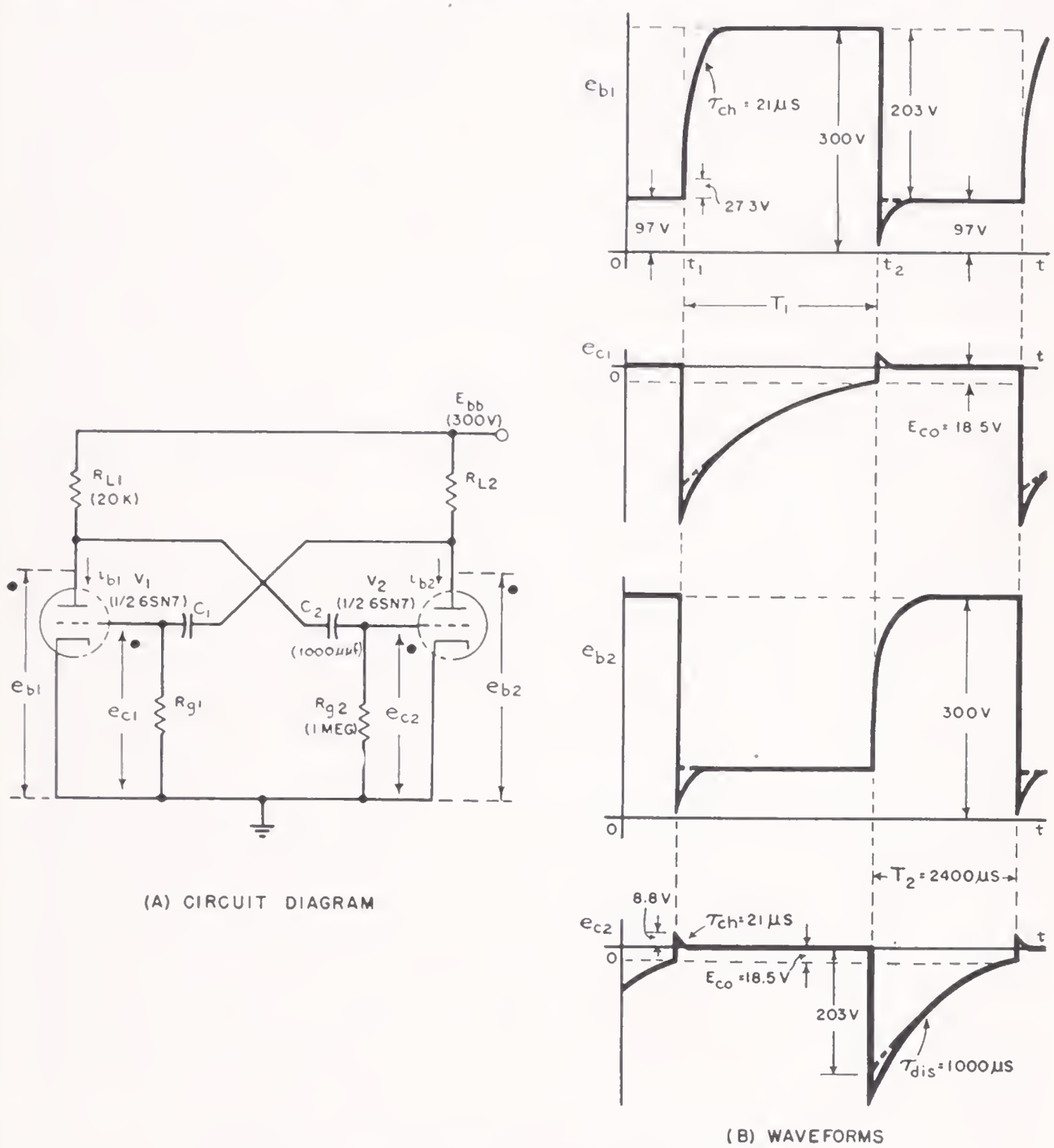


FIG. 28. Free-running plate-coupled multivibrator.

two halves of the circuit are alike, the tube currents may at first be nearly equal. However, balanced currents cannot be maintained; any slight change in either current will bring about a cumulative change of the currents as follows: Suppose a slight increase in the current drawn by tube V_2 occurs. This increase causes an increase of the voltage drop across the resistor R_{L2}, and thus a decrease in e_{b2}. Because the voltage across capacitor C_1 cannot change instantly, the decrease in e_{b2} is accompanied

by a decrease in e_{c1}. A decrease of e_{c1} reduces the plate current in V_1. Thus the increase in the plate current of V_2 must be accompanied by a decrease in the current of tube V_1. In the same manner, the decrease of i_{b1} brings about an increase of i_{b2}. Thus the slight initial increase of i_{b2} sets up a cumulative, or *regenerative, switching action* which ends with i_{b1} reduced to zero and i_{b2} increased to a maximum value. Though described as if it occurred slowly, this switching actually occurs extremely rapidly—in a small fraction of a microsecond in well-designed multivibrators.

In order to cut off plate current in V_1, the grid of that tube must be driven negative beyond the cutoff voltage. The negative grid voltage results from a charge on the capacitor C_1. Because this charge leaks off through resistor R_{g1}, the grid voltage does not remain negative indefinitely but rises toward zero as the capacitor discharges. As soon as the cutoff voltage is reached, plate current begins to flow in tube V_1 and a second switching action takes place. This switching action is like the first, except that now i_{b1} is increasing and i_{b2} decreasing. Therefore it ends with V_1 carrying a maximum current and V_2 cut off; that is, during the switching action the current is suddenly transferred from the plate circuit of one tube to that of the other.

Since tube V_2 also is cut off because of the voltage across a capacitor, it cannot remain so indefinitely. As soon as the charge on C_2 has decreased so that the grid voltage of V_2 reaches cutoff, a third switching action occurs, and current is suddenly transferred from V_1 to V_2. Thus the cycle of operation of a multivibrator consists of periods during which first one and then the other tube is cut off, the switching from one condition to the other being very rapid.

The waveforms of voltage obtained are shown in Fig. 28*B*. The plate voltage of either tube is low when that tube is conducting and approaches the plate-supply voltage when the tube is cut off. The grid voltage of either tube is forced suddenly negative when the plate voltage of the other tube drops, and then it rises exponentially toward zero as the coupling capacitor discharges. When the cutoff voltage is reached, switching occurs, the tube conducts, and the grid voltage jumps suddenly positive as far as grid current will allow.

Details of the waveforms may be explained in terms of an example, using the circuit constants indicated in parentheses in Fig. 28*A*.

At time t_1 (see Fig. 28*B*), tube V_1 is cut off suddenly. The resulting plate voltage e_{b1} and grid voltage e_{c2} can be found exactly as was done for the *R-C* peaker in Art. 7. The equivalent circuit for the charge of C_2 during the cutoff interval of V_1 is shown in Fig. 29*A*. The initial voltage across the capacitor is the voltage to which the capacitor discharged in the preceding period when the tube was conducting. The voltage at the plate side of C_2 immediately before V_1 is cut off may be found by drawing

a 20-K load line from 300 volts on the 6SN7 curves and using the $e_c = 0$ point. The value found is 97 volts. The corresponding voltage at the grid side of C_2 is the cutoff voltage for V_2, which is -18.5 volts. The initial capacitor voltage is therefore $97 + 18.5 = 115.5$ volts.

The initial current in the equivalent circuit and initial values of e_{c2} and e_{b1} may be calculated exactly as for the R-C peaker. The results, 8.8 volts for e_{c2} and $97 + 27.3$ volts for e_{b1}, are indicated on the waveforms of Fig. 28B. Both the grid voltage e_{c2} and the plate voltage e_{b1} rise suddenly by $8.8 + 18.5 = 27.3$ volts when V_1 is cut off; as C_2 charges, e_{b1} continues to rise exponentially toward 300 volts and e_{c2} decreases toward zero. The time constant of the exponentials, obtained from the equivalent circuit, is 21 μs.

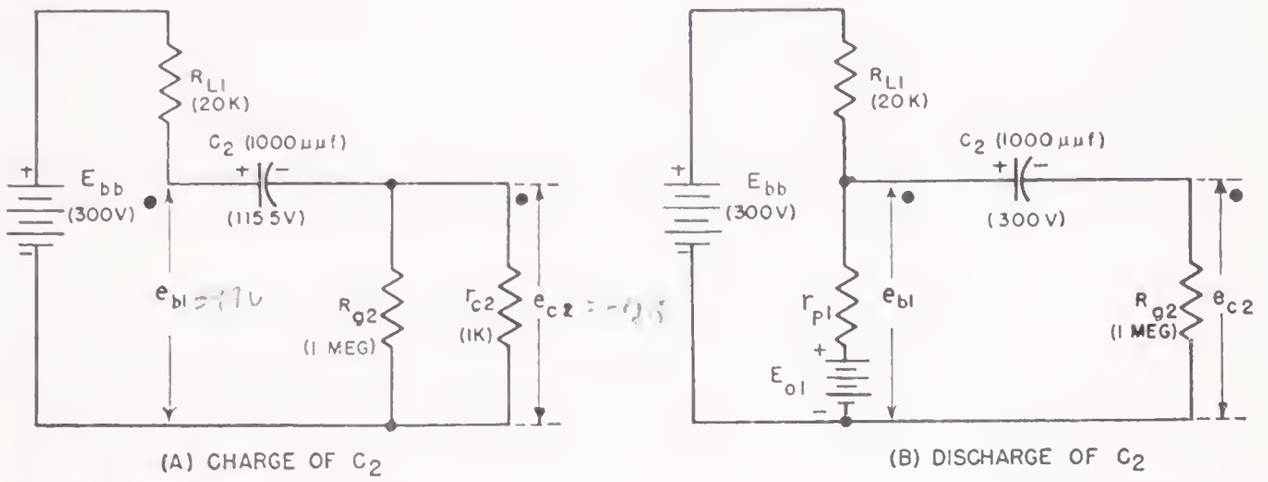


FIG. 29. Equivalent circuits for coupling between plate of V_1 and grid of V_2 in Fig. 28A.

Before time t_2 is reached, e_{b1} and e_{c2} have attained their final values. Thus C_2 is charged to 300 volts, and the equivalent circuit that applies in the period following time t_2, when V_1 is conducting, is that of Fig. 29B. The value of E_{o1} is actually variable during this period because e_{c1} is first driven a few volts positive and then decreases to zero in the same way that e_{c2} was driven positive and decreased in the preceding period. Nevertheless, as a first approximation, e_{c1} is assumed to be constant at zero when V_1 conducts; and therefore in the Thévenin's theorem equivalent circuit for Fig. 29B, the portion of the circuit to the left of the e_{b1} terminals is replaced by a battery of 97 volts in series with the parallel combination of $R_{L1} = 20$ K and $r_{p1} = 7.0$ K.

Waveforms may be calculated from this Thévenin-theorem circuit, but the results obtained are negligibly different from those obtained if e_{b1} is assumed constant at 97 volts throughout the conducting period of V_1. Because R_{g2} is so large relative to r_{p1} and R_{L1} in parallel, the maximum voltage across r_{p1} and R_{L1} is 1 volt. This small voltage is neglected in the waveforms of Fig. 28B, the dotted line in the e_{b1} wave is at a constant voltage of 97 volts, and the initial point on the dotted curve for e_{c2} is

below zero by 300 − 97 = 203 volts—the sudden jumps in the e_{b1} and e_{c2} waveforms are equal.

After dropping suddenly to −203 volts, the grid voltage e_{c2} rises exponentially toward zero as C_2 discharges. In computing the time constant of this rise, the parallel combination of r_{p1} and R_{L1} is neglected relative to R_{g2} for consistency with the previous approximation, and the time constant is found to be τ_{dis} = 1,000 $\mu\mu$f × 1 megohm = 1,000 μs, as indicated in Fig. 28*B*. The exponential rise of e_{c2} continues until the cutoff voltage (−18.5 volts) is reached; then the next switching process occurs.

The positive peak in the wave of e_{c1}, neglected in calculating values on the dotted waves, causes a negative peak in the wave of e_{b1}. The negative peak is an amplified and inverted replica of the peak in the e_{c1} wave, as shown by the solid line in Fig. 28*B*. Because of the long time constant of discharge of C_2, this peak appears also in the grid-voltage wave e_{c2} superimposed upon the exponential curve already determined as shown by the solid line. The actual magnitudes of these peaks can be found from the plate characteristics of the tube, but it is seldom necessary to do so.

The waveforms of e_{b1} and e_{c2} have now been determined by considering the charge and discharge of capacitor C_2. In a similar manner, the charge and discharge of capacitor C_1 could be considered and the waveforms of e_{b2} and e_{c1} determined. These calculations are omitted here because they are the same as those already made. If the two halves of the multivibrator are exactly alike, the multivibrator is called *balanced* or *symmetrical*. Then the two plate-voltage waves are alike, except for a shift along the time axis, as are also the two grid-voltage waves.

The periods T_1 and T_2 (during which tubes V_1 and V_2, respectively, are cutoff) are the most often needed constants of a multivibrator. The time T_2 depends on the discharge time constant of C_2, but *the cutoff time is not equal to the time constant.* From the exponential part of the grid-voltage curve, the cutoff time may be calculated. Thus in Fig. 28*B*, the *dotted* curve for e_{c2} is an exponential, has a time constant of 1,000 μs, and rises from −203 to −18.5 volts in the interval T_2. Therefore

$$18.5 = 203\epsilon^{-T_2/1,000}$$

$$\epsilon^{-T_2/1,000} = \frac{18.5}{203} = 0.0912$$

$$T_2 = 1,000 \times 2.4 = 2400\ \mu\text{s}$$

Observe that the cutoff time depends not only upon the time constant but also upon the ratio of the voltages involved. In this example, it is 2.4 times as long as the time constant.

From the constants of the coupling circuit including C_1, the period T_1 could be calculated in an exactly similar fashion. The sum of T_1 and T_2 is the repetition period of the multivibrator. Thus the circuit constants determine the frequency at which the free-running multivibrator operates.

14. Driven Plate-coupled Multivibrator. A different multivibrator is the circuit of tubes V_1 and V_2 in Fig. 30*A*. Tube V_3 is a trigger tube used to apply a sharp pulse of voltage to initiate each cycle of multivibrator operation. The trigger tube remains cutoff and has no effect upon the multivibrator, except during the trigger pulse.

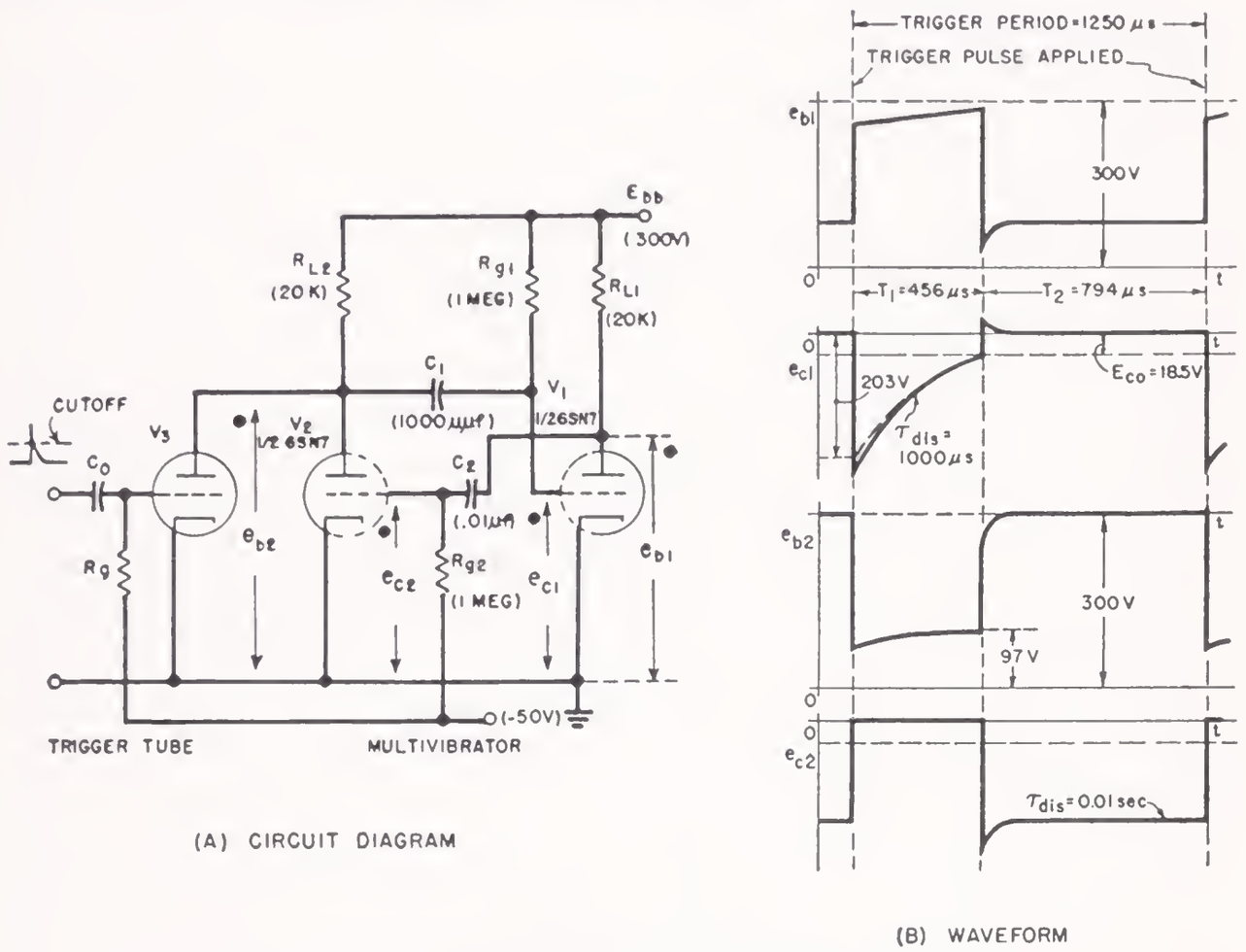


FIG. 30. One-cycle plate-coupled multivibrator.

The multivibrator of Fig. 30*A* cannot operate without a driving pulse. Because the grid of V_2 is returned to a negative voltage below cutoff, V_2 would remain permanently cutoff and V_1 permanently conducting in the absence of trigger pulses. V_2 is said to be the *normally off* tube, and V_1 the *normally on* tube.

When a narrow positive pulse is applied to the grid of V_3, the plate voltage of V_3 and V_2 is made to drop momentarily. This negative pulse is applied by C_1 to the grid of V_1 and causes a positive pulse of voltage at the plate. The positive pulse is applied to the grid of V_2 by C_2, and if its amplitude is sufficiently large, it drives the grid voltage of V_2 momentarily above cutoff. Once tube V_2 begins to draw current, a regenerative

switching process occurs; V_2 becomes the conducting tube and V_1 the cutoff tube. This switching process is identical with that in the free-running multivibrator.

Tube V_1 cannot remain cutoff permanently, however, because its grid is returned to +300 volts. As capacitor C_1 discharges, the grid voltage of tube V_1 rises exponentially toward +300 volts. As soon as the cutoff voltage is reached, V_1 begins to conduct again, a second switching process occurs, and the normal condition (V_2 off, V_1 on) is reestablished. This condition persists until another trigger pulse is applied because, even with C_2 discharged to a steady-state voltage, the grid voltage of V_2 is −50 volts, and this value maintains V_2 cutoff. Thus the circuit is a one-cycle multivibrator.

To trigger this multivibrator, a *negative* pulse may be applied to the plate of V_2 (as in Fig. 30*A*) or directly to the grid of V_1, or a positive pulse may be applied to the plate of V_1 or the grid of V_2. Whatever circuit arrangement is used, *the positive pulse at the grid of the cutoff tube actually causes the cycle of operation to start.* Unless this pulse is *large enough to cause current to begin to flow in the cutoff tube,* triggering is not accomplished. Note that triggering can occur even though the trigger pulse alone is not sufficient to cut off the conducting tube V_1.

Typical values of circuit elements are indicated in parentheses in Fig. 30*A*, and corresponding waveforms are drawn in Fig. 30*B*. Details of these waveforms may be determined exactly as was done for the free-running multivibrator. The waveform of e_{b1} differs somewhat from the voltage at a plate of a free-running multivibrator (compare Fig. 28*B*, p. 93) in that the vertical rise of e_{b1} is of much greater magnitude. This effect is obtained because the large negative bias for V_2 and the long discharge time constant of C_2 (0.01 sec) ensure that e_{c2} can rise only very slightly while V_2 is cutoff. Thus C_2 may be considered a coupling capacitor, and e_{b1} and e_{c2} are similar to voltages in coupling circuits (compare Fig. 20*B*, p. 80). Consequently the grid of V_2 is not driven appreciably positive, and the negative peak at the bottom of the e_{b2} wave is not pronounced

The cutoff time of V_1 is the duration of the pulses produced by the multivibrator and is indicated as T_1 in Fig. 30*B*. Because of the positive-grid return for V_1, this interval is shorter than the interval T_2 calculated for the free-running multivibrator, even though the initial and final values of grid voltage and the time constant of the exponential are the same. To calculate this interval, note that the exponential in the e_{c1} wave is rising toward +300 volts. Therefore, the initial value of the exponential, *measured with respect to its steady-state value* is 300 + 203 = 503 volts. Similarly, the value of the exponential when V_1 begins to conduct is 300 + 18.5 = 318.5 volts. The cutoff interval

may then be calculated from the equation

$$318.5 = 503\epsilon^{-T_1/1,000}$$

which yields

$$T_1 = 1{,}000 \times 0.456 = 456\ \mu s$$

The length of time that V_2 is cutoff (T_2 in Fig. 30*B*) is determined by the period of the applied trigger pulse, because V_2 once cutoff remains so until the next pulse is applied. Since the trigger period is assumed to be 1,250 μs, T_2 is $1{,}250 - 456 = 794$ μs.

Either positive or negative pulses having a duration of 456 μs and timed by the trigger pulses may be obtained from this circuit. If positive pulses are desired, the output is obtained from either the plate of V_1 or the grid of V_2. If negative pulses are desired, the plate of V_2 is used as the output terminal.

Another driven-multivibrator circuit is obtained if the grid of V_2 in Fig. 30*A* is returned through R_{g2} to the cathode. Such a circuit is capable of free running, but at a very low frequency because of the long discharge time constant of C_2. It is useful as a driven multivibrator because of this long time constant; if trigger pulses are applied before the discharge of C_2 allows the grid voltage of V_2 to reach cutoff, the multivibrator operates in synchronism with the driving source. If the discharge of C_2 is made very slow and a large-amplitude trigger pulse is used, the possibility of the multivibrator switching too soon or too late is avoided; moreover, a very nearly rectangular pulse is obtained.

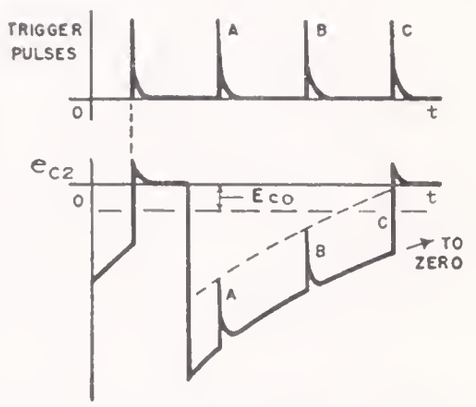


FIG. 31. Grid-voltage waveform for frequency division.

If this same circuit, with the grid of V_2 returned to cathode, is used with trigger pulses of reduced amplitude, frequency division may be obtained. The waveform of e_{c2} for a 3:1 division ratio is drawn in Fig. 31. The positive trigger pulse at the grid of V_2 is made so small that the first two times it is applied, at *A* and *B* in Fig. 31, the grid voltage is not raised to the cutoff value. By the time of the third trigger pulse at *C*, the grid voltage has risen enough to allow the trigger pulse to extend above cutoff and to trip the circuit.

15. Cathode-coupled Multivibrator. The circuit diagram of a much used cathode-coupled multivibrator is drawn in Fig. 32*A*. Note that *R-C* coupling is provided from the plate of V_1 to the grid of V_2, exactly as in the plate-coupled multivibrator, but the coupling from plate of V_2 back to grid of V_1 is omitted. In its place, a cathode resistor common to the two tubes is provided. Thus tube V_2 operates as a cathode follower—

changes in its grid voltage are repeated at its cathode and affect the grid-to-cathode voltage of V_1.

This circuit is a one-cycle multivibrator because, in the absence of trigger pulses, tube V_2 (having the positive grid return) conducts and raises the voltage of both cathodes. The grid voltage of V_1 is adjusted to a fairly low value by the choice of R_1 and R_2, and therefore the cathode of V_1 is

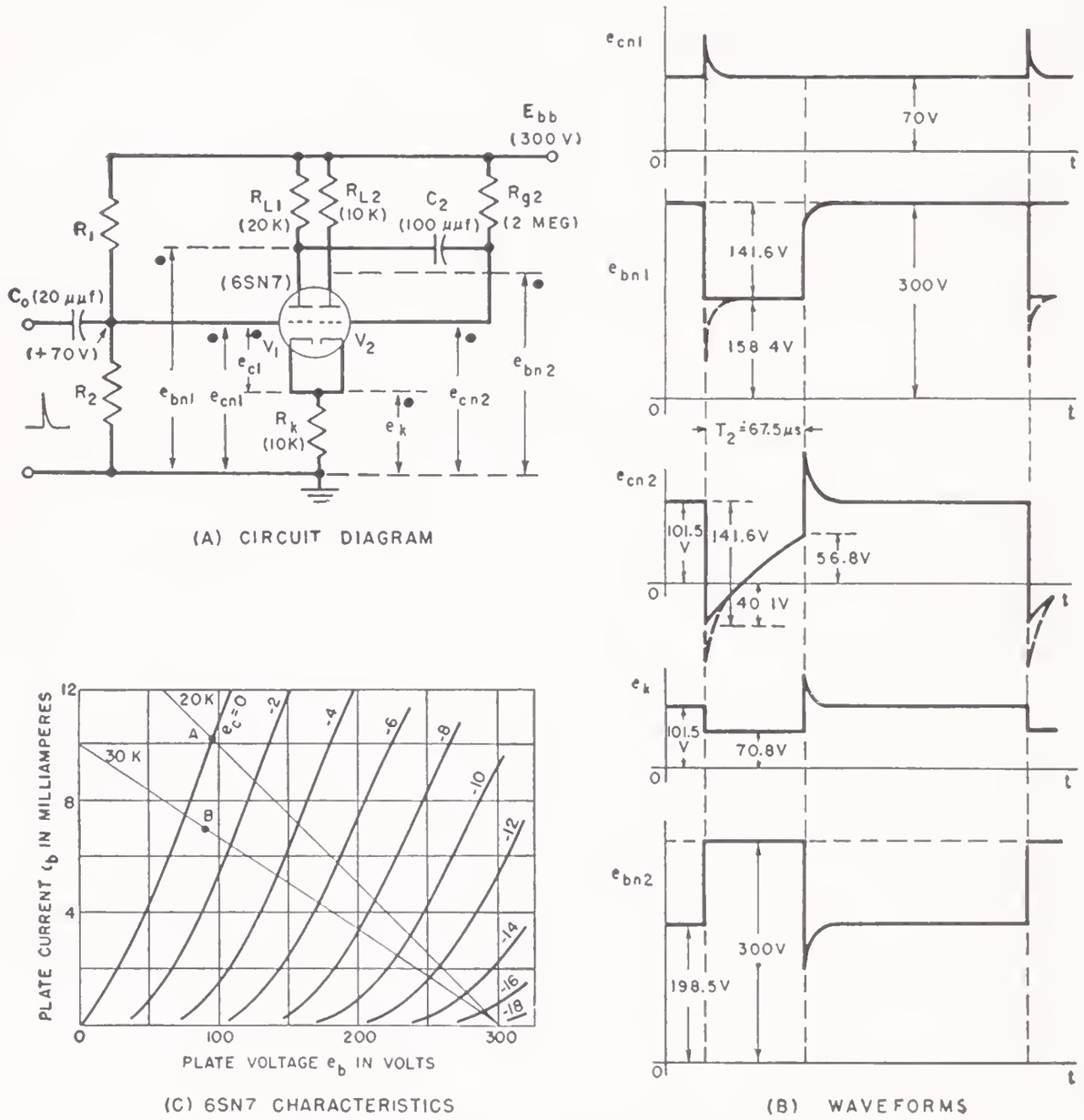


FIG. 32. Cathode-coupled multivibrator.

positive with respect to the grid by a large enough voltage to keep V_1 cutoff.

The multivibrator may be driven by positive pulses applied to the grid of the normally off tube V_1, as in Fig. 32*A*. If the pulse raises the grid voltage sufficiently to start current in V_1, a switching process occurs and the plate current changes from V_2 to V_1. The switching proceeds as follows: Current in V_1 causes the plate-to-ground voltage e_{bn1} to drop. Because of capacitor C_2, the grid-to-ground voltage e_{cn2} drops equally; and since V_2 is connected as a cathode follower, e_k also drops. The

decrease in e_k means an increase in the grid-to-cathode voltage for V_1 and an increase in the current of V_1.*

Once cut off V_2 remains so, while C_2 discharges, until the cutoff point is reached and plate current begins to flow. Then the second switching process occurs. The current in V_2 raises the cathode voltage enough to cut off V_1, and the high plate voltage of V_1 helps to turn on V_2.

Figure 32*B* shows the waveforms of voltages *with respect to ground*. These are the voltages most conveniently observed with an oscilloscope. The output voltage is the plate voltage of V_2 with respect to ground if a positive pulse is desired or the plate voltage of V_1 with respect to ground if a negative pulse is wanted. Details of these waveforms, using the circuit constants in Fig. 32*A*, are determined in the following example.

The grid-to-ground voltage of V_1 is assumed to be set to +70 volts by the voltage divider composed of resistors R_1 and R_2. This value is maintained except during the sharp positive pulses applied through C_o, as shown in Fig. 32*B*. Before the application of a pulse, V_1 is cutoff and therefore e_{bn1} is +300 volts. Because of the large value of R_{g2} (2 megohms) the grid-to-cathode voltage of V_2 may be assumed to be zero. Therefore, point *A* on the 20-K load line for V_2 in Fig. 32*C* may be located; from this point the plate current of V_2 is found to be 10.15 ma. Thus e_k and e_{cn2} are each equal to $10 \times 10.15 = 101.5$ volts before the application of a trigger pulse, and e_{bn2} is equal to

$$300 - 10 \times 10.15 = 198.5 \text{ volts}$$

A check should be made to see that V_1 is actually cut off as assumed. The plate-to-cathode voltage is $300 - 101.5 = 198.5$ volts, and the corresponding grid-to-cathode voltage for cutoff is -12 volts. The actual grid voltage is $70 - 101.5 = -31.5$ volts. Therefore the assumption of cutoff is correct, and a trigger pulse having an amplitude of at least 19.5 volts is required to drive the multivibrator.

At the time of the trigger pulse, V_2 is cut off and V_1 begins to conduct. The 30-K load line in Fig. 32*C* may be used to determine the current in V_1. (Note that R_{g2} is so large that the effect on V_1 of current in C_2 and R_{g2} is negligible.) Because of the cathode resistor, the operating point (point *B* in Fig. 32*C*) is *not* at $e_c = 0$, but rather at a point where $e_{cn1} = e_k + e_{c1}$ has the given value of 70 volts, which point may be

* It might seem that the increase of current in V_1 would offset the decrease of current in V_2 and result in no change of the cathode voltage and no switching process. This, however, is not true. With constant grid voltage for V_1 and decreasing grid voltage for V_2, the total current in R_k must decrease, and therefore e_k decreases. The changes of current in V_1 cause the cathode follower V_2 to operate into a very low load resistance; thus the changes of e_k are considerably less than the corresponding changes of e_{cn2}.

located by trial-and-error calculations.[1] From point B the current in V_1 is found to be 7.08 ma; and therefore during the conducting period of V_1, e_k is 70.8 volts and e_{bn1} is $300 - 20 \times 7.08 = 158.4$ volts. The grid voltage e_{cn2} must drop suddenly by the same amount as e_{bn1}, and therefore it reaches the initial value of $101.5 - 20 \times 7.08 = -40.1$ volts.

The grid-to-cathode cutoff voltage of V_2 for this period is -14 volts, corresponding to a plate-to-cathode voltage of $300 - 70.8 = 229.2$ volts. Thus V_2 again conducts when e_{cn2} rises to $70.8 - 14 = 56.8$ volts. All these values of voltages are indicated on the waveforms in Fig. 32B.

When V_1 is first cut off at the end of the multivibrator pulse, a positive peak appears in the grid-voltage wave e_{cn2} and the plate voltage e_{bn1} rises exponentially toward 300 volts exactly as e_{b1} does in plate-coupled multivibrators.[2] If the trigger pulse is of large amplitude, the negative pulses (shown dotted in Fig. 32B) appear in the waveforms of e_{bn1} and e_{cn2} because a large trigger pulse increases the current in V_1 momentarily to a higher value than the circuit can maintain.

To determine the duration of the multivibrator pulse, which is the cutoff time for V_2, the time constant of the exponential rise of e_{cn2} is needed. Because R_{g2} is much larger than R_{L1} in parallel with the plate-to-ground resistance of V_1,* this time constant is very nearly

$$100\ \mu\mu\text{f} \times 2\ \text{megohms} = 200\ \mu\text{s}$$

The initial amplitude of the exponential in the e_{cn2} waveform, measured from the steady-state value of $+300$ volts, is 340.1 volts. The value at the cutoff point is $300 - 56.8 = 243.2$ volts. From the exponential relation

$$243.2 = 340.1\epsilon^{-T_2/200}$$

The value of T_2, the cutoff interval of V_2, is found to be 67.5 μs.

The positive pulse applied to the grid of V_1 is not the only method of triggering a cathode-coupled multivibrator. Negative pulses at the plate of V_1 or grid of V_2 may be used equally well. Often a trigger tube is connected in parallel with V_1, and positive pulses applied to the grid of the trigger tube cause negative pulses at the plate of V_1. If an available positive trigger pulse is followed by a negative overshoot, a trigger

[1] See Art. 3, p. 61, for similar calculations applied to the cathode follower.

[2] In calculating the amplitude and time constant of this peak, it must be remembered that the rise of grid voltage causes the cathode voltage to rise. The equivalent grid-to-ground circuit for V_2 is the input circuit of a cathode follower (Fig. 7, p. 66).

* The plate-to-ground resistance is very high because of the cathode resistor. Any increase of current in V_1 increases the magnitude of the negative grid-to-cathode voltage because of the voltage drop across R_k; therefore the plate-to-cathode voltage of V_1 must increase greatly to allow the increases of current. The apparent resistance from plate-to-ground is $r_p + R_k(\mu + 1)$.

tube may be used to clip away the overshoot, which otherwise might cause cutoff of V_1 immediately after the positive trigger started conduction in V_1.

A particular advantage of this circuit is that the duration of the output pulse may be controlled by adjustment of the bias voltage applied to the grid of V_1. In the example of Fig. 32, if R_1 or R_2 is changed so that e_{cn1} is less than 70 volts, the current in V_1 during the multivibrator pulse is decreased. As a result, the plate voltage of V_1 is higher during the pulse and the cathode voltage lower. The exponential rise of e_{cn2} therefore begins at a higher value and ends at a lower value, and the cutoff time of V_2 is decreased. The multivibrator pulse is made shorter by decreasing the voltage at the grid of V_1 and longer by raising this voltage.

The voltage e_{cn1} must, of course, remain within a certain range if the multivibrator is to operate correctly. As e_{cn1} is decreased, the values of e_{cn2} at the beginning and end of the pulse approach each other. At the particular value of e_{cn1} corresponding to zero pulse length, these two values become equal, and for this or lower values of e_{cn1} the multivibrator cannot be triggered. If e_{cn1} is increased too far, a point is reached at which V_1 cannot be held off between pulses, and free-running operation is obtained. At the end of each pulse, V_1 is cut off momentarily by the narrow positive peak in the e_k waveform (see Fig. 32*B*). As soon as this positive peak is over, V_1 is allowed to conduct and another pulse is begun. High values of e_{cn1} may also cause V_1 to draw grid current during the pulse. Grid current in V_1 interferes with proper control of e_{cn1}.

By correct choice of R_{L1} and R_k, *the duration of the output pulse can be made a very nearly linear function of the voltage* e_{cn1} for variations within the range of correct operation. This property is useful in range-measuring circuits. For example, a step or other range-marker waveform can be produced at the time of the trailing edge of the multivibrator pulse, and the marker can be varied in position by adjustment of e_{cn1}. The voltage e_{cn1} can then be used to indicate the range to the marker with considerable accuracy.

To understand how the linear relation can be obtained, consider first the relationship between the grid voltage e_{cn1} and the change of e_{cn2} during a pulse. The value of e_{cn2} at the beginning of the pulse is a linear function of the plate current of V_1 during the pulse. The value of e_{cn2} at the end of the pulse also is essentially a linear function of i_{b1} because, to the extent that the cutoff voltage of V_2 is proportional to e_{b2}, the value of e_{cn2} is a linear function of $e_k = i_{b1}R_k$. The plate current i_{b1} during the pulse is very nearly a linear function of e_{cn1}. Thus, except for tube nonlinearities, a linear relationship between e_{cn1} and the change of e_{cn2} during a pulse exists.

The change of e_{cn2} during a pulse is proportional to the pulse duration,

provided that the average discharge current for C_2 is independent of pulse duration. The discharge current varies considerably *during a pulse;* in the example of Fig. 32, it decreases from (340.1)/2 = 170 μa at the beginning to (243.2)/2 = 121.6 μa at the end. Nevertheless, R_{L1} and R_k can be so adjusted that decreasing e_{cn1} raises the grid voltage of V_2 at the beginning of the pulse and simultaneously lowers the grid voltage at the end in such a way that the *average* discharge current for C_2 is nearly the same for any length of pulse. Therefore proper adjustment of these resistors makes the pulse duration a very nearly linear function of e_{cn1}.

16. Phantastron Delay Circuit. Another circuit for producing a controllable delay time is the phantastron of Fig. 33*A*. This circuit employs a pentagrid tube, Type 6SA7, instead of a double triode, yet its operation is very similar to that of the cathode-coupled multivibrator.

The type 6SA7 tube may be considered equivalent to a pentode with the screen grid replaced by the structure comprising grids 2, 3, and 4. Grids 2 and 4 are connected together and to a positive potential and function as a screen grid. Grid 3 functions as an additional control grid; some of the electrons pass through it and reach the plate, but others are returned toward the cathode and eventually collected by the screen grid. Changing the voltage of grid 3 has two effects: (1) The division of current between plate and screen grid is altered—a negative increment of grid 3 voltage *reduces* the plate current and *increases* the screen-grid current. (2) The cathode current is changed—some of the returned electrons pass through the screen grid and tend to decrease the electric field near the cathode, thus causing the decrease of plate current to be greater than the corresponding increase of screen-grid current. Grid 1, like the control grid in a pentode, controls the sum of the plate and screen-grid currents with little effect upon the division ratio. Grid 5 functions as a suppressor grid. Grids 1 and 3, like the control grid of a triode, draw current when at a positive voltage with respect to the cathode.

Figure 33*A* is drawn in such a way as to show the very close relationship between the phantastron circuit and the cathode-coupled multivibrator of Fig. 32*A*. The plate, grid 3, and cathode of the phantastron tube are analogous to the elements of triode V_1 in Fig. 32*A*. The screen grid, grid 1, and cathode of the phantastron correspond to the elements of triode V_2.

The plate section may be thought of as a normally off "tube," and the screen-grid section as a normally on "tube." In fact, before a trigger pulse is applied, grid 1 is at cathode potential because of the positive return of R_g and a large screen-grid current flows. As a result, the cathode voltage e_k is well above the voltage of grid 3 and plate current is very nearly cutoff. Application of a positive trigger pulse to grid 3 starts the flow of plate current, thereby reducing the plate-to-ground

voltage e_{bn} and the grid 1 voltage e_{cn1}. As a result, the cathode voltage drops as in a cathode follower and an increase of plate current is caused by the higher grid-3-to-cathode voltage. Thus a switching process starts exactly as in the cathode-coupled multivibrator.

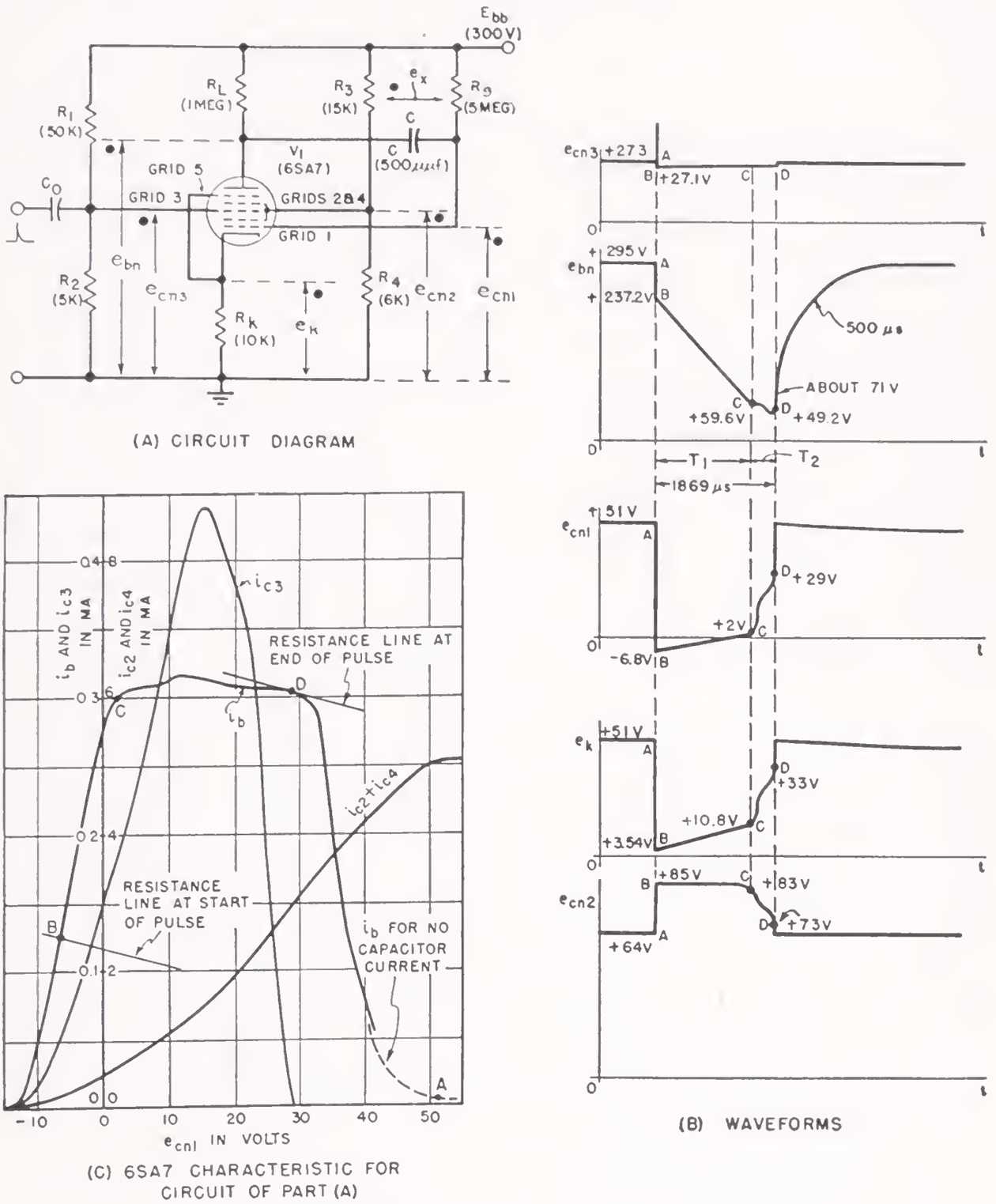


FIG. 33. Basic phantastron delay circuit.

In the cathode-coupled multivibrator, this switching process ends with V_2 cutoff. In the phantastron, it cannot go so far because, if e_{cn1} were reduced enough to cut off screen-grid current, plate current also would be stopped. Without plate current, the drop of e_{cn1} could not occur.

To show more clearly the limit of the switching process and other

details of the waveforms produced, the characteristics of Fig. 33*C* are used. These curves apply only for the circuit constants in Fig. 33*A*; they are obtained by applying various constant voltages to grid 1 and measuring the corresponding values of plate current i_b, screen-grid current $i_{c2} + i_{c4}$, and grid 3 current i_{c3}. During the determination of these characteristics, an adjustable resistor in series with a microammeter is connected across R_L to furnish a plate-current component equal to the current that flows through C in normal operation. Operating conditions when e_{cn1} is less than e_k are to be studied, and therefore current in C equals the current in R_g, which is $(E_{bb} - e_{cn1})/R_g$. Hence, at each point for which data are taken, the adjustable resistor is set to draw this amount of current; for example, at $e_{cn1} = 30$ volts, the resistor is adjusted for a current of $(300 - 30)/5 = 54$ μa. For use in determining the initial state of the circuit, a small portion of the i_b curve (shown dashed in Fig. 33*C*) is determined with no allowance for capacitor current. Corresponding portions of similar $i_{c2} + i_{c4}$ and i_{c3} curves are indistinguishable from the solid-line curves.

Figure 33*C* indicates that no electrons leave the 6SA7 cathode for values of e_{cn1} below -15 volts. An increase of e_{cn1} above this value causes the expected smooth increase of screen-grid current. Small increases of e_{cn1} also cause increases of plate current and grid 3 current. With further increase of e_{cn1}, these currents pass through maxima and begin to decrease because of the positive cathode voltage resulting from the current (mostly $i_{c2} + i_{c4}$) in R_k. For values of e_{cn1} above 30 volts, the grid-3-to-cathode voltage is negative and i_{c3} is zero. For values of e_{cn1} above 55 volts, the negative grid-3-to-cathode voltage is of sufficient magnitude to prevent any electrons from passing through the third grid toward the plate.

The curves of Fig. 33*C* may be applied to determine points on the waveforms of Fig. 33*B*. In the steady-state condition before the trigger pulse is applied (point *A* of Fig. 33*B*), grid-3 current is cutoff, plate current is nearly cutoff, capacitor current is zero, and $e_{cn1} = e_k$ because grid 1 is returned to E_{bb}. The exact operating point (*A* in Fig. 33*C*) is determined by a trial-and-error calculation for which the final trial is $e_{cn1} = 51$ volts, $i_{c3} = 0$, $i_{c1} = (300 - 51)/5{,}000 = 0.05$ ma,

$$i_{c2} + i_{c4} = 5.06 \text{ ma}$$

$i_b = 0.005$ ma (obtained from the dashed curve), and

$$e_k = 10(0.05 + 5.06 + 0.005) = 51.1 \text{ volts}$$

which is essentially equal to the value assumed for e_{cn1}. The plate voltage is

$$e_{bn} = 300 - 0.005 \times 1{,}000 = 295 \text{ volts}$$

and the grid 3 voltage is

$$e_{cn3} = 300 \times \frac{5}{5 + 50} = 27.3 \text{ volts}$$

By Thévenin's theorem, the screen-grid-supply voltage divider is equivalent to an 85.7-volt battery in series with a 4.28-K resistor. Therefore

$$e_{cn2} = 85.7 - 4.28 \times 5.06 = 64.0 \text{ volts}$$

These initial voltages are marked in Fig. 33*B*.

Application of a positive trigger pulse starts the switching process by causing an increase of i_b and a decrease of e_{bn} and e_{cn1}. According to Fig. 33*C*, the decrease of e_{cn1} causes an additional increase of i_b, and the switching process continues. After the peak of the i_b curve is passed, an equilibrium point marking the end of the switching process is reached at values of i_b and e_{cn1} that satisfy not only the i_b, e_{cn1} relation of the tube but also an i_b, e_{cn1} relation determined by the circuit elements connected to the tube. The end point of the switching process can be found by a trial-and-error calculation, but a quicker method is to locate the intersection of an appropriate resistance line with the i_b, e_{cn1} curve of Fig. 33*C*. The resistance line specifies the relation between i_b and e_{cn1} fixed by resistors R_L and R_g and the voltage on capacitor C.

The resistance line may be determined by expressing i_b as the sum of the currents in R_L and R_g. Thus, if e_x is the capacitor voltage,

$$i_b = \frac{300 - e_x - e_{cn1}}{10^6} + \frac{300 - e_{cn1}}{5 \times 10^6}$$

$$10^6 i_b = (360 - e_x) - {}^6{}_5 e_{cn1} \tag{19}$$

At the time of the switching process, e_x is 295 − 51 = 244 volts, and

$$10^6 i_b = 116 - {}^6\!/_5 e_{cn1} \tag{20}$$

Equation 20 describes the start-of-pulse resistance line in Fig. 33*C* for which $i_b = 116$ μa at $e_{cn1} = 0$ and the slope is $-{}^6\!/_5$ μa per volt.

From the intersection B between this line and the i_b curve, the values $e_{cn1} = -6.8$ volts and $i_b = 124$ μa are read. Corresponding points on the other curves indicate $i_{c2} + i_{c4} = 0.18$ ma and $i_{c3} = 50$ μa. Voltages e_{cn2} and e_{cn3} can be found from these current values with the help of Thévenin's theorem. The plate voltage is

$$e_{bn} = e_x + e_{cn1} = 244 - 6.8 = 237.2 \text{ volts}$$

and e_k is the voltage produced across R_k by the cathode current

$$i_k = 0.18 + 0.124 + 0.05 = 0.354 \text{ ma}$$

The initial pulse voltages are indicated at points B in Fig. 33*B*.

After the first instant of the pulse, e_x decreases because of the capacitor-discharge current. Decrease of e_x moves the resistance line in Fig. 33*C* upward (see Eq. 19), and the operating point on the i_b curve therefore moves upward from point *B*. The phantastron continues in the plate-section-conducting condition until the resistance line reaches the end-of-pulse position in Fig. 33*C*, tangent to the i_b curve at point *D*. Further decrease of e_x moves the resistance line so that it does not cross the i_b curve at any point, and the plate-section-conducting condition can no longer be maintained. The second switching process therefore occurs, the pulse ends, and the phantastron returns to its original condition.

Voltage waveforms during the pulse may be determined point by point using several positions of the resistance line. To illustrate the procedure, calculations are made at the end point *D* and on intermediate point *C* as follows: Between points *B* and *C* in Fig. 33*C*, i_b rises sharply, causing a considerable decrease of e_{bn}, but e_{cn1} increases only a few volts (see the waveforms of Fig. 33*B*). The decrease of e_x during this interval is therefore mostly associated with the decrease of e_{bn}. At point *C*, e_{cn1} is 2 volts and i_b is 0.3 ma. From Eq. 19, e_x is 57.6 volts, and therefore e_{bn} is $57.6 + 2 = 59.6$ volts. Between points *C* and *D*, i_b remains nearly constant, the change of e_{bn} is not great, and e_{cn1} increases to allow for the continued decrease of e_x. Thus the rounded corner of the e_{cn1} wave is produced. At point *D*, e_{cn1} is 29 volts, i_{bn} is 0.305 ma, and therefore e_x is 20.2 volts and e_{bn} is 49.2 volts.

The duration of the phantastron pulse is the sum of time intervals T_1 and T_2 marked in Fig. 33*B*. During interval T_1, the capacitor current decreases from $(300 + 6.8)/5 = 61.33$ μa to $(300 - 2)/5 = 59.6$ μa. The change of current is very small, and therefore the capacitor voltage changes at an essentially constant rate, and the portion of the e_{bn} waveform between *B* and *C* is very nearly a straight line. With small error, the discharge current may be assumed to have a constant value

$$\frac{61.33 + 59.6}{2} = 60.5 \ \mu\text{a}$$

The change of capacitor voltage in the T_1 interval is

$$244 - 57.6 = 186.4 \text{ volts}$$

and thus the capacitor relation $i = C de/dt$ yields

$$T_1 = \frac{C\Delta e}{i} = \frac{500 \times 186.4}{60.5} = 1{,}540 \ \mu\text{s}$$

Similarly, the change of voltage during interval T_2 is 37.4 volts, and the discharge current varies between 59.6 and 54.2 μa. The approximation of a constant current (56.9 μa) is slightly less accurate than before, but

the time interval is short so that small percentage errors in its calculation are unimportant. Thus

$$T_2 = \frac{500 \times 37.4}{56.9} = 329 \ \mu s$$

The pulse duration is therefore 1,540 + 329, or 1,869 μs.

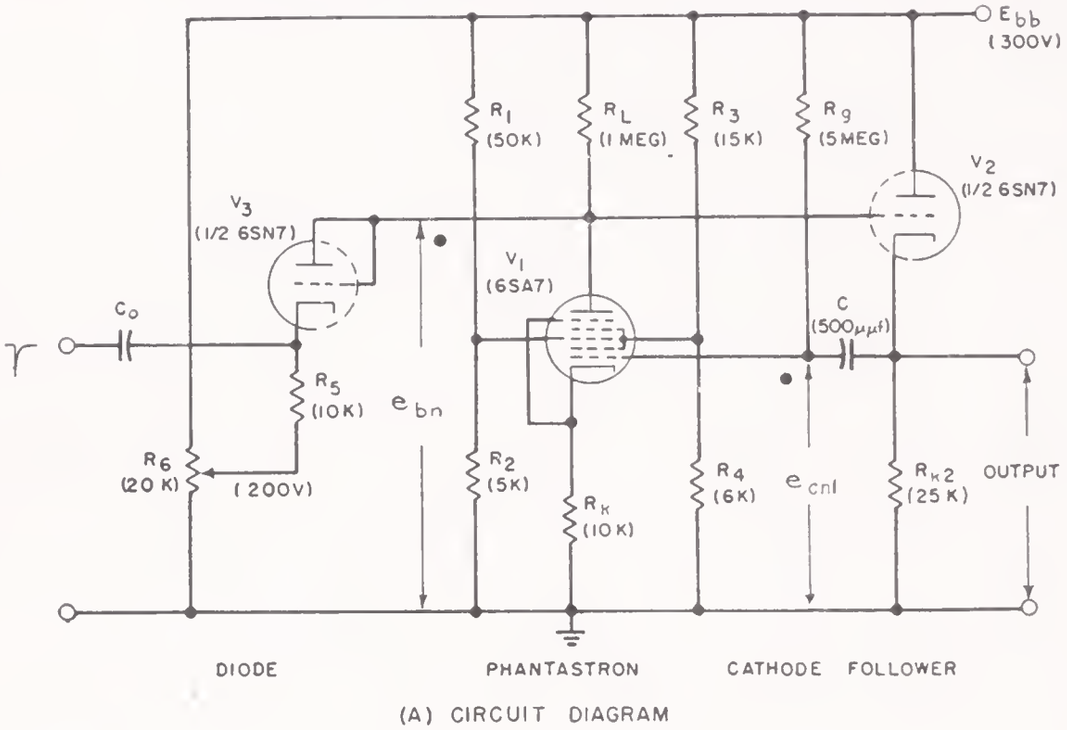
The very high value of R_L (1 megohm) is used in order that the change of e_{cn1} during the pulse may be small and the rate of discharge of C nearly constant. The constant discharge rate is needed to make the pulse duration vary in proportion to changes of a control voltage. A disadvantage of the high value of R_L is that after the pulse the plate voltage returns relatively slowly to +300 volts, as shown in Fig. 33*B*. Because R_L is high, e_{cn1} is limited effectively to its value at the grid-current point, and the plate voltage can jump initially only as far as e_{cn1} does, which is a little more than 22 volts. Following the jump, e_{bn} rises exponentially with a time constant of 500 μs as C charges through R_L (the equivalent grid resistance is negligible compared with R_L). Because of this relatively slow recharging of C, a long period must elapse after the pulse before the circuit is ready for a second triggering.

If a cathode follower is added to the basic phantastron circuit, as in Fig. 34*A*, the time required for recharging capacitor C between pulses is greatly reduced. The plate circuit of the pentagrid tube V_1 has merely to raise the grid voltage of the follower; current from the plate-cathode circuit of the follower is used to charge the capacitor. During the pulse the cathode follower causes no essential change in the circuit operation, though the loading effect of R_g on the plate circuit is removed, and the changes of e_{cn1} are a little less than the corresponding change of e_{bn}. Waveforms of e_{bn} and e_{cn1} appear in Fig. 34*B* for comparison with the waveforms in Fig. 33*B*.

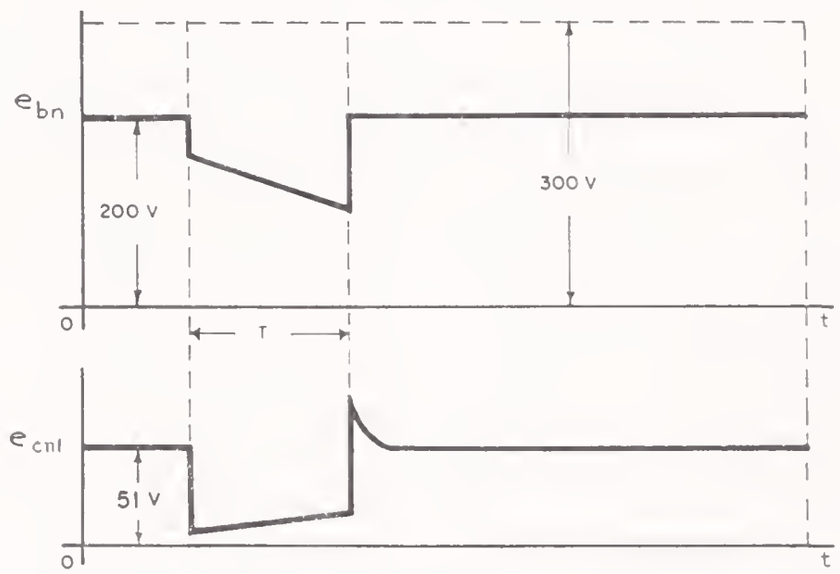
In Fig. 34*A* a diode for controlling the pulse duration is shown also, and a negative trigger pulse is applied through the diode to the plate of V_1. The diode limits the voltage to which C is charged between pulses. Thus if potentiometer R_6 is set for +200 volts at the tap, the diode conducts when the plate voltage of V_1 rises to this value. Between pulses e_{bn} is limited very nearly to 200 volts, because R_5 and R_6 have very low resistances in comparison with the 1-megohm resistor R_L. Corresponding to e_{bn} = 200 volts, the cathode-to-ground voltage of V_2 is 201 volts. Since the grid voltage e_{cn1} is not affected by the diode, the pulse is begun with only 201 − 51 = 150 volts across C instead of the 244 volts obtained without the diode. Because the diode is cutoff during the pulse, the voltage across C at the end of the pulse is in no way affected by the setting of R_6. Thus the change of voltage on C during the pulse is controlled directly by the adjustment of R_6. Since C is discharged at an essentially

constant rate, the duration of the pulse varies in an almost perfectly linear fashion with the control voltage at the tap on R_6.

The output voltage is normally taken from the cathode of V_2, as indicated in Fig. 34*A*. The waveform of this voltage is essentially the same as that of e_{bn} shown in the figure. The trailing edge of the pulse,



(A) CIRCUIT DIAGRAM



(B) WAVEFORMS

FIG. 34. Phantastron with cathode follower and diode.

which is delayed by an amount controlled by R_6, does not rise with extreme rapidity because the shunt capacitance of the circuit must be charged through the high resistance R_L. Consequently, if a delayed trigger pulse is desired, the output waveform is generally used to trigger a blocking oscillator (see Art. 21) in order to obtain a sharp pulse.

17. Tube and Wiring Capacitances. The capacitances between electrodes in tubes and from circuit elements and wiring to ground,

though neglected in the preceding articles, actually have an appreciable effect upon the waveforms obtained in timing circuits. In sine-wave amplifiers these small capacitances reduce the gain at high frequencies. In timing circuits they tend to round off sharp corners of pulses, to delay steep wavefronts slightly, and to broaden very narrow pulses and reduce their amplitudes. These effects occur because a quick change of a voltage requires an equally quick charge and discharge of the capacitances. Since the sharp corners and steep sides of a pulse give rise to the high-frequency terms in its Fourier series (Art. 3, Chap. VI), the effect of capacitance on sudden changes corresponds to the attenuation of high frequencies.

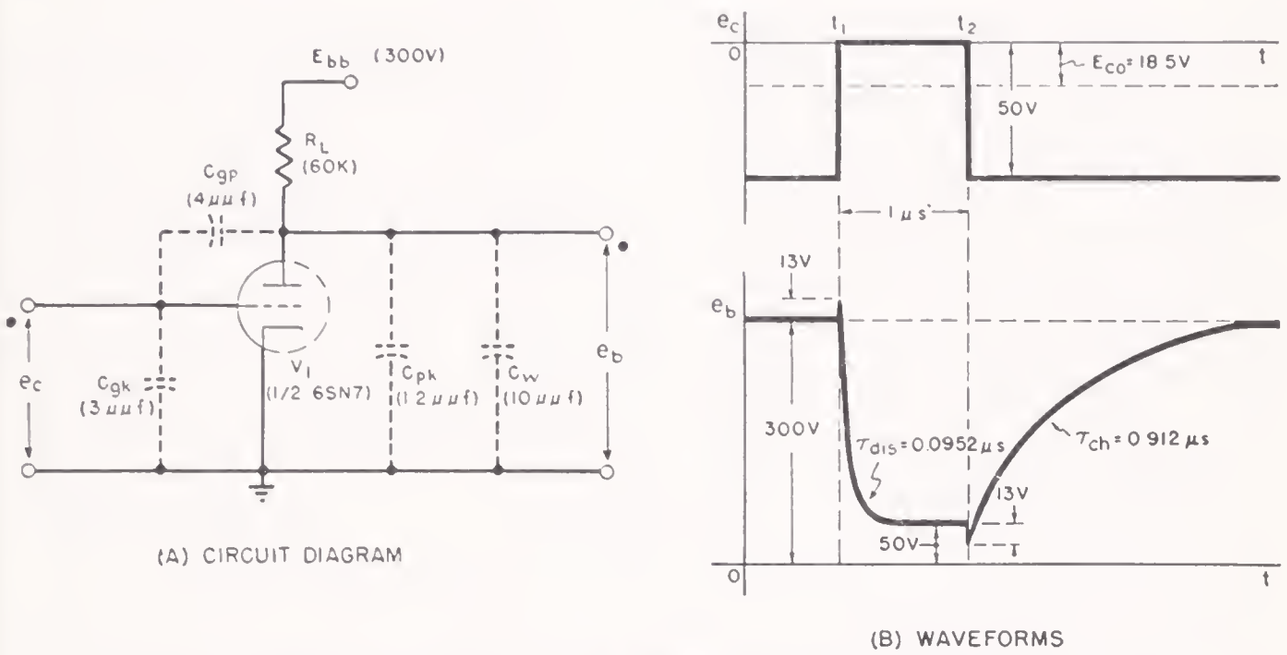


FIG. 35. Effect of tube and wiring capacitances on overdriven amplifier.

An example of the effects of shunt capacitance is provided by the pulse amplifier of Fig. 35*A*. Suppose that the circuit constants are as indicated and that the driving source is capable of providing the input voltage wave e_c (see Fig. 35*B*) with essentially vertical sides, despite the input capacitance of the amplifier. Interelectrode capacitances of the tube, together with an assumed value of wiring capacitance, are indicated on the figure.

Immediately before time t_1, the tube is held cutoff by the negative grid voltage, and e_b is equal to the supply voltage, 300 volts. Therefore the equivalent circuit before time t_1 is that of Fig. 36*A* with switches S_1 and S_2 open. Voltages across the capacitors are as indicated. The jump in e_c at time t_1 corresponds to closing S_1 and S_2 simultaneously—the 50-volt jump is applied through C_{gp} to the plate circuit, and simultaneously the tube is made to conduct. Closing S_1 completes a series circuit through the two capacitors; the net voltage of 50 volts transfers charge from C_{gp} to C_{pk} and C_w at a rate limited only by the resistance of

the source of e_c. The charge on C_{gp} before time t_1 is

$$350 \times 4 = 1{,}400 \ \mu\mu\text{coulombs}$$

and that on $C_{pk} + C_w$ is $300 \times 11.2 = 3{,}360$ $\mu\mu$coulombs. Very shortly after S_1 is closed, this total charge is shared by the capacitors in parallel. Thus the initial value of e_b during the pulse is

$$\text{Initial } e_b = \frac{1{,}400 + 3{,}360}{4 + 11.2} = 313 \text{ volts}$$

and the equivalent circuit applicable during the pulse is that shown in Fig. 36*B*.

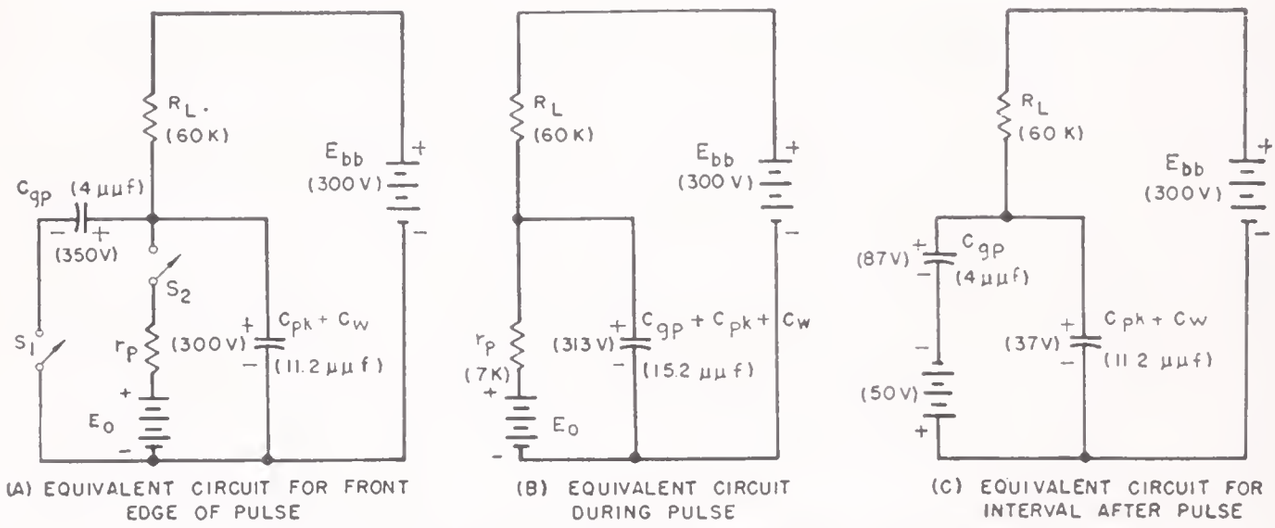


FIG. 36. Equivalent circuits for amplifier of Fig. 35*A*.

Another way to explain the rise of e_b at the front edge of the pulse is to note that capacitors C_{gp} and $C_{pk} + C_w$ form a capacitance voltage divider for the application of a fraction of the *change* of e_c to the plate circuit as a *change* of e_b. Because the wavefront of e_c is so steep, the resistor currents are negligible in comparison with capacitor currents. Capacitance voltage dividers function with *changes* of voltage in the same way as resistance voltage dividers do with *total* voltages, thus[1]

$$e_b = e_c \times \frac{1/(C_{pk} + C_w)}{\dfrac{1}{C_{gp}} + \dfrac{1}{C_{pk} + C_w}} = 50 \times \frac{1/11.2}{\dfrac{1}{4} + \dfrac{1}{11.2}} = 13.1 \text{ volts}$$

which agrees with the result obtained in terms of total charge.

Following the time t_1 in Fig. 35*B*, capacitors C_{gp}, C_{pk}, and C_w in parallel discharge exponentially to 50 volts (the value obtained from a 60-K load line drawn on the 6SN7 curves). The value of r_p over the region of the $e_c = 0$ line between $e_b = 50$ volts and $e_b = 313$ volts is about 7 K. There-

[1] Since $R = e/i$ and $1/C = (de/dt)/i$, the reciprocal of capacitance is analogous to resistance in the voltage-divider relation.

fore, the time constant of the discharge is

$$\tau_{dis} = 15.2 \times \frac{60 \times 7}{60 + 7} \times 10^{-3} = 0.0952 \ \mu s$$

as indicated in Fig. 35*B*.

When time t_2 is reached, the capacitance voltage divider causes e_b to drop 13 volts when e_c drops 50 volts. Following this instant, the equivalent circuit of Fig. 36*C* applies, and e_b returns exponentially to 300 volts. From the figure, the time constant is

$$\tau_{ch} = 15.2 \times 60 \times 10^{-3} = 0.912 \ \mu s$$

The entire waveform of e_b is drawn in Fig. 35*B*. Note that the negative-going edge at the front of the e_b pulse is much steeper than the positive-going edge at the end of the pulse, because the plate-circuit resistance is low when the tube conducts.

From this example, it is clear that tube and wiring capacitances must be minimized if waveforms with sharp corners and steep edges are to be produced. Pentodes are frequently used to obtain steep edges and especially to reduce the effect of grid-to-plate capacitance. To determine the effect of pentode capacitances, consider C_{gp} to be zero, and use the values given in tube handbooks for C_i and C_o in place of C_{gk} and C_{pk}, respectively.

Reducing the load resistor is *not* the equivalent of reducing the capacitances, even though in linear amplifiers the resistance and capacitance have equivalent effects upon high-frequency response. In overdriven circuits the output wave is frequently applied to another overdriven tube to change the grid voltage of the second tube from zero to far below cutoff. Then the only concern is to make steep the small portion of the wavefront that corresponds to the region between zero and cutoff for the second tube; time constants are not of direct importance. Reducing the load resistance always reduces the time constants but may either reduce or increase the steepness of particular parts of the wavefronts. In the amplifier circuit of Fig. 35*A*, for example, the initial slope of a negative-going jump in the output wave is unaffected by the size of the load resistance, for at the top point the tube is discharging the capacitances and the load resistance is carrying no current. At lower points on a negative-going edge, the wavefront is made steeper if R_L is made larger because the voltage toward which the exponential is falling is made smaller. For the positive-going edge, however, lower values of R_L increase the steepness at each point, as well as reducing the time constant (though little is gained in steepness if R_L is reduced to values less than r_p).

Capacitances are important in the operation of cathode followers as well as amplifiers. *When both are operated in the linear range of their characteristics,* cathode followers can produce steeper wavefronts than amplifiers because of the low internal resistance of the follower. *When both are overdriven,* cathode followers are *not* superior to amplifiers in generating steep edges. If a positive-going wavefront is applied as an input voltage first to an amplifier and second to a cathode follower comprising the same tube and load resistor, the changes of output voltage will be equally rapid in the two circuits, provided each grid is driven to zero voltage (or to equal positive voltages) and provided capacitances are the same in each circuit. This is true because the same current is available in each circuit to charge the capacitances. Similarly, if a large negative-going wavefront is applied, both tubes will be cut off and the changes of output voltage will be equally rapid. Actually, the cathode follower has the disadvantages of possessing extra capacitance from cathode to heater and of requiring a greater rate of change of input voltage to produce the same rate of change of output voltage.[1] Note that a cathode follower can produce steep *rises* of output voltage because the tube conducts to cause the rise, but not so steep *drops* of output voltage. (Contrast this statement to the waveform for the amplifier in Fig. 35*B*.)

C. *R-L-C* CIRCUITS

Inductors as well as resistors and capacitors are employed in radar timing circuits. Self-inductors are used to generate narrow pulses of voltage or, in combination with capacitors, to produce timing oscillations. Mutual inductors (usually closely coupled iron-core transformers) are useful as interstage coupling devices and in special feedback circuits. Inductors and capacitors connected together to form artificial transmission lines serve to control time relations between waveforms and to produce pulses of special shape.

The operation of *R-L* circuits is analogous to that of the *R-C* circuits

[1] For very sudden changes of input voltage, the input capacitance of an amplifier is C_{gk} in parallel with the series combination of C_{gp} and $C_{pk} + C_w$ (see Fig. 35*A*), because in the first instant resistor currents are negligible compared with capacitor currents. The corresponding capacitance for a cathode follower is C_{gp} in parallel with the series combination of C_{gk} and $C_{pk} + C_w + C_{kh}$, where C_{kh} is the capacitance between cathode and heater. These two input capacitances are nearly the same in value; therefore the higher rate of change of input voltage required by a cathode follower actually is a disadvantage.

The approximation used with linear amplifiers that $C_i = C_{gk} + C_{gp}(1 + A)$ is valid only for input-voltage changes so slow that the tube and wiring capacitances have little effect upon the output voltage. The equivalent approximation for the cathode follower is $C_i = C_{gp} + C_{gk}(1 - A)$.

in Sec. B. The following facts about inductance correspond to those stated for capacitance in Art. 4 and should be memorized:

1. The voltage-current relation for an inductor is

$$e = L\frac{di}{dt} \tag{21}$$

2. The current through an inductor cannot be changed instantly.
3. The average voltage across an inductor is zero.[1]
4. Transients in single-inductor R-L circuits are always exponential, with time constant L/R.

Inductors in timing circuits most frequently appear as part of R-L-C circuits rather than the simpler R-L circuits, because even if an actual capacitor is not placed in the circuit, the capacitance between turns of the coil is generally large enough to be important. The combination of elements most often encountered is resistance, inductance, and capacitance in parallel.

18. Transients in Parallel R-L-C Circuits. Corresponding to the time constant associated with R-C or R-L circuits, there are two constants that characterize parallel R-L-C circuits. These constants are

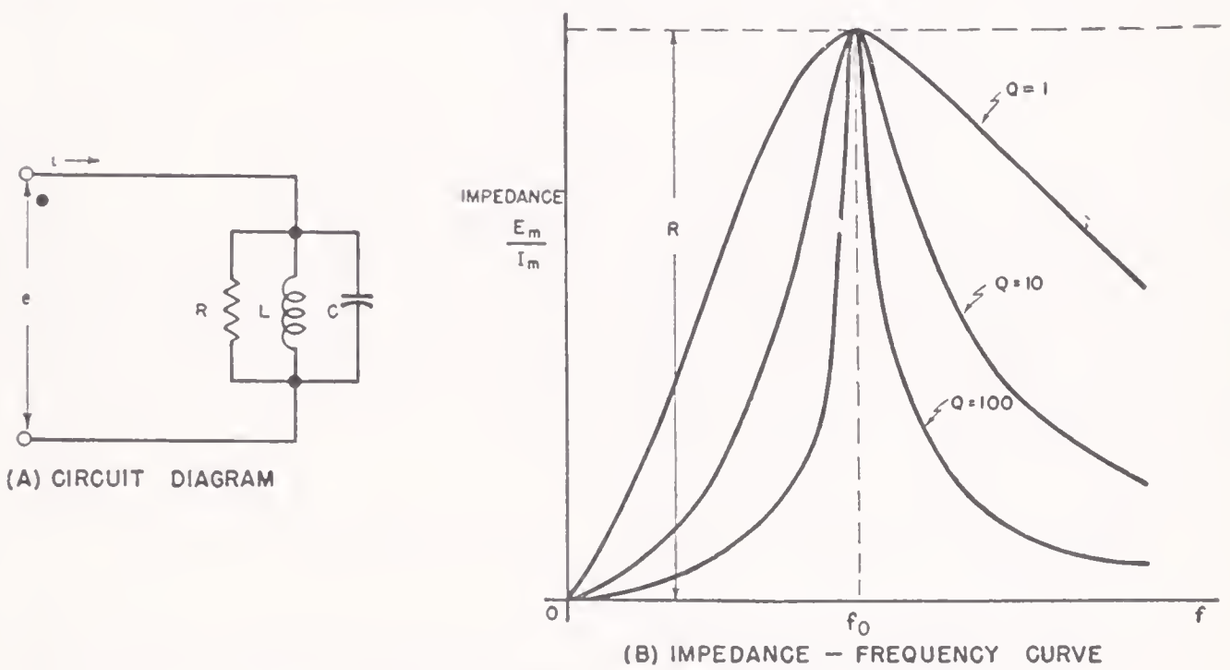


FIG. 37. Parallel R-L-C circuit.

the *resonant frequency* f_0 and the quality factor Q. Their meaning is most readily explained by reference to the circuit and steady-state impedance-frequency curves of Fig. 37. A sinusoidal voltage,

$$e = E_m \sin \omega t$$

[1] Statement 3 assumes that the current is the same at the beginning and end of the averaging period and that the inductor has no winding resistance.

where ω is 2π times the frequency, is assumed to be applied to the circuit; as a result, a sinusoidal current, $i = I_m \sin(\omega t + \phi)$, flows from the source of e. The magnitude of the *impedance*, or ratio E_m/I_m, is plotted in Fig. 37*B* as a function of the frequency of the sine waves.

The resonant frequency f_0 is the frequency at which the 90-deg leading current in the capacitor is exactly equal in magnitude to the 90-deg lagging current in the inductor. At this frequency the net current drawn from the source is the current in R, and the impedance of the entire circuit equals R. At lower frequencies, the inductor draws a greater current than the capacitor; hence the total current from the source is increased, and the impedance is less than R. Similarly, at higher frequencies, the capacitor current exceeds the inductor current, and the impedance is decreased. Therefore, f_0 marks the peak of the impedance-frequency curve. The value of f_0 can be determined from the requirement of equality of the currents in the capacitor and inductor. At resonance

$$E_m \times 2\pi f_0 C = \frac{E_m}{2\pi f_0 L}$$

and therefore

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{22}$$

or

$$T_0 = 2\pi\sqrt{LC} \tag{23}$$

where T_0 is the period of a resonant-frequency wave.

The quality factor Q specifies the suddenness of the decrease of impedance above and below the resonant frequency. This factor is defined in Art. 2 of Chap. IX in terms of power loss and stored energy for any form of resonator, and the general definition is shown in that article to reduce to

$$Q = \omega_0 CR = \frac{R}{\omega_0 L} = \frac{R}{\sqrt{L/C}} \qquad \text{for parallel } R\text{-}L\text{-}C \text{ circuit} \tag{24}$$

Each of these expressions is the ratio of the resistance in the parallel circuit to the reactance at resonance of either the inductor or the capacitor. Since the same voltage is applied to the resistance and the reactance of a parallel circuit, an equivalent expression is

$$Q = \left.\frac{\text{current in either } L \text{ or } C}{\text{current in } R}\right|_{\substack{\text{at}\\ \text{resonance}}} \qquad \text{for parallel } R\text{-}L\text{-}C \text{ circuit} \tag{25}$$

If at resonance the leading current in C and the equal lagging current in L are very large relative to the current in R, then a small shift from the resonant frequency causes a considerable increase in source current and a

large decrease of impedance. If, on the other hand, a small capacitor current is balanced by a small inductor current at resonance, the change of impedance with frequency is not so rapid. In this way, a high Q corresponds to a sharply resonant circuit, and a low Q to a broad maximum of the impedance curve, as indicated in Fig. 37*B*.

Frequently a definition of Q that is the reciprocal of one of the expressions of Eq. 24 is encountered. Such a value for Q is obtained if the basic definition of Chap. IX is applied to a *series* R-L-C circuit instead of the parallel circuit here considered. Thus

$$Q = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 CR} = \frac{\sqrt{L/C}}{R} \qquad \text{for series } R\text{-}L\text{-}C \text{ circuit} \qquad (26)$$

The expressions for Q in Eq. 26 may be applied also to a circuit comprising a capacitance C in parallel with the *series* combination of R and L.

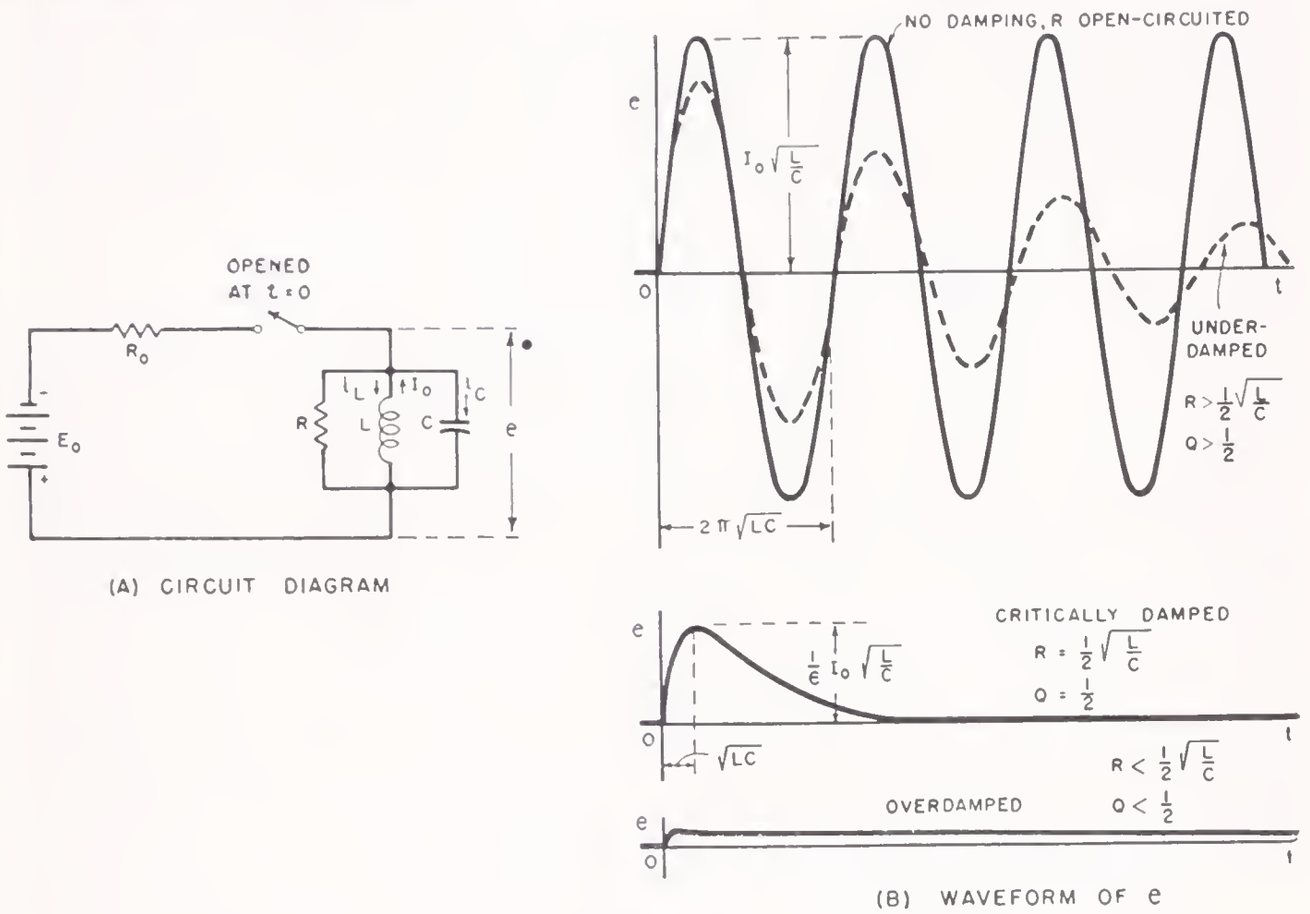


FIG. 38. Voltage transients in parallel R-L-C circuits.

Transients in R-L-C circuits are not all of the same shape. Instead, the form of the transients depends upon the Q of the circuit. In Fig. 38*A*, an initial steady current I_0 is assumed to be built up in the inductance L, and the variations of e after the switch is opened are plotted in Fig. 38*B*. The curves are all drawn for the same values of L and C but for different values of Q corresponding to different values of R.

If R is open-circuited, so that Q is infinite, the transient is a sinusoidal voltage wave, as in Fig. 38*B*. Because the inductor current cannot

change instantly when the switch is opened, the current I_0 flows from L into C and causes the voltage e to increase because of the charge placed on C. The voltage across the capacitor causes the current to decrease. The voltage reaches its maximum value when the current is zero, because a further decrease of current means a reversal of direction and a discharge of the capacitor. At $t = 0$, energy is stored in the inductance because of the magnetic field set up by the current I_0. When the inductor current decreases to zero and the capacitor voltage is at its maximum, this energy is all transferred to the electric field of the capacitor. As the capacitor voltage decreases after reaching its peak, the energy is returned to the inductance, the return of energy being complete when $e = 0$. The reversed current in L then charges C in the direction to make e negative; and so the process continues, energy passing from one element to the other and back and e oscillating above and below the zero axis. Since the circuit is assumed to be lossless, the amount of energy transferred does not change and the oscillation continues at constant amplitude. In fact, the energy in the electric field of the capacitor at each peak of the oscillation equals the initial energy in the magnetic field of the inductor, or

$$\tfrac{1}{2}Ce_{\max}^2 = \tfrac{1}{2}LI_0^2$$

and therefore

$$e_{\max} = I_0\sqrt{\frac{L}{C}} \tag{27}$$

where $e_{\max}$ is the maximum value of e.

The sinusoidal form of the oscillation can be verified as follows: The current i_C (see Fig. 38*A*) flowing into C is related to e by the equation

$$i_C = C\frac{de}{dt}$$

The voltage e is also the voltage across L and is related to the current i_L (see Fig. 38*A*) in the inductor by

$$e = L\frac{di_L}{dt}$$

With the switch open and R open-circuited, $i_L = -i_C$, and therefore

$$e = -L\frac{di_C}{dt} = -LC\frac{d^2e}{dt^2}$$

or

$$\frac{d^2e}{dt^2} = -\frac{1}{LC}e \tag{28}$$

Equation 28 states that the rate of change of the *slope* of the curve of e is proportional in magnitude and opposite in sign to the ordinate of the

curve. A sine wave possesses this same property, as may be shown by successive differentiation:

$$\frac{d^2}{dt^2}(E_m \sin \omega t) = \frac{d}{dt}(\omega E_m \cos \omega t) = -\omega^2 E_m \sin \omega t \tag{29}$$

where ω is the angular frequency $2\pi f$. Comparing Eqs. 28 and 29 shows that, for the constants of proportionality to be equal, the angular frequency of the wave of e must be $1/\sqrt{LC}$. That is, *the frequency of oscillations in the lossless L-C circuit equals the resonant frequency* f_0.

If a high resistance R is used in the circuit of Fig. 38*A*, oscillating variations of e are still obtained but the amplitude of oscillation decreases from one cycle to the next, as shown by the dotted curve in Fig. 38*B*. The oscillation is said to be *damped*. Energy initially in the inductor is only partly transferred to the capacitor during the first quarter cycle; a small part of the energy is lost as heat in the resistor. Additional loss of energy occurs during each cycle, and the envelope of the oscillations decreases exponentially, because the loss of energy during each half cycle is proportional to the stored energy at the beginning of the half cycle. The frequency of the damped oscillations can be shown to be

$$f = f_0 \sqrt{1 - \left(\frac{1}{2Q}\right)^2} \tag{30}$$

and is slightly less than the resonant frequency of the circuit. The difference between f and f_0 is negligible except for very low values of Q—if Q is 5, the difference is only 0.5 per cent.

If the value of R is reduced, the rate of decay of the oscillations is increased. As long as the variations of e remain oscillatory, the circuit is said to be *underdamped*. When, however, R is reduced to $\frac{1}{2}\sqrt{L/C}$ so that $Q = \frac{1}{2}$, the damping is so rapid that no negative swing of e is obtained and the circuit is said to be *critically damped*. The critically damped transient of e is drawn in Fig. 38*B*. Note that the peak value of e is much less than the peak with no damping, because much of the energy initially in the inductance is lost to the resistor before e reaches its peak.

If the value of R is further reduced so that $Q < \frac{1}{2}$, the *overdamped* transient in Fig. 38*B* is obtained. This transient is the sum of two exponentials. A short-time-constant exponential of the form $1 - \epsilon^{-t/\tau}$ accounts for the quick rise of e immediately after $t = 0$, and a long-time-constant exponential of the form $\epsilon^{-t/\tau}$ accounts for the slow return to zero. As R is reduced in value, the short time constant becomes shorter, the long time constant longer, and the peak value of e lower.

For very low values of Q, the overdamped response may be approximated as follows: During a short period following $t = 0$, the inductor

is a source of a constant current I_0. This current charges C to the voltage I_0R along an exponential curve having the very short time constant RC. After reaching the value I_0R, the voltage e remains nearly constant but decreases slowly as the current in the inductor decreases. The decay of e is exponential and has the time constant L/R, because the change is so slow that the capacitor current is negligible. If the overdamping is extreme, therefore, the response approaches that of the R-L circuit, the capacitor merely preventing a perfectly vertical rise of e at $t = 0$.

In actual circuits C may be a capacitor or merely the capacitance between windings of the inductor and the shunt capacitance of the circuit. A resistor may be placed in parallel with the tuned circuit if critically damped or overdamped response is desired, or the damping may be only that caused by coil losses if a highly oscillatory transient is wanted. Not all the losses in the coil are exactly equivalent to those produced by a parallel resistance. For example, the resistance of the inductor winding is in series with the inductance. Losses in the inductor, however, are generally small, and any resonant circuit with small losses is approximated to a high degree of accuracy by a parallel R-L-C circuit of equal Q. For convenience, therefore, all losses may be represented by an equivalent shunt resistor.

19. Ringing Circuit. One method of range measurement employed in radar sets uses a resonant circuit in which a transient oscillation is begun at the time of each transmitted pulse. The oscillation is allowed to continue for the duration of the range sweep and is damped out before the next transmitted pulse. In this way the range sweep is divided into a series of known intervals determined by the period of the oscillation.

A common circuit for producing such oscillations is drawn in Fig. 39*A*. The name *ringing circuit* is applied to it because of the analogy between the oscillations it produces and mechanical oscillations that cause a ringing sound. Before each transmitted pulse, tube V_1 is conducting with $e_c = 0$, and a steady current flows through the tube and inductance. At the time of a transmitted pulse, time t_1, the tube is cut off by a negative gate pulse applied to its grid. The inductor current therefore flows into the capacitor, charging it in the direction to make e_k negative, and starting the oscillation drawn in Fig. 39*B*. Losses in the L-C circuit are made as small as possible, with the result that the amplitude of oscillation is nearly constant throughout the duration of the input gate pulse. At the end of the gate pulse, time t_2, the tube again conducts and a second oscillation is started. However, the conducting tube is equivalent to a resistance shunted across the L-C circuit and causes the oscillations to be damped quickly.

Typical values of circuit constants are indicated in Fig. 39*A*. Before

the time t_1, the current in the tube and inductance is 14.2 ma, this value being read from the 6SN7 characteristics for $e_c = 0$ and $e_b = 125$ volts. This current causes a negligible voltage drop in the winding resistance of L; therefore e_k is zero and, corresponding to $e_c = 0$, e_{cn} is zero. Because damping of the oscillation during the cutoff period of the tube is slight, the period of oscillation and the amplitude of the first negative peak of

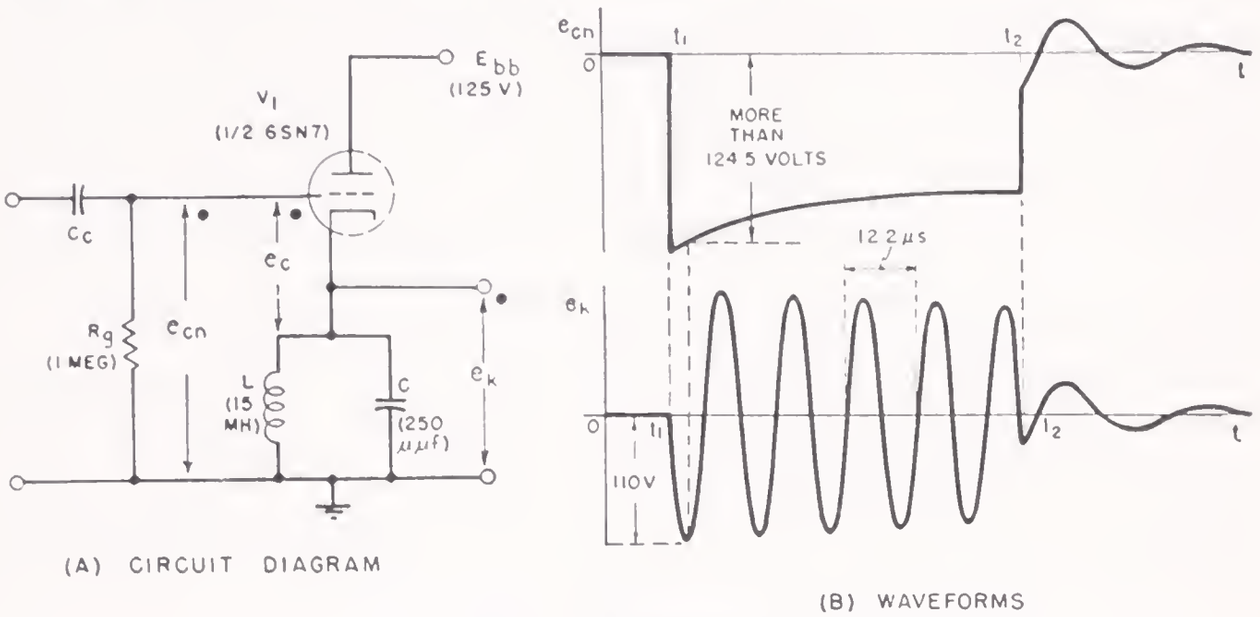


FIG. 39. Ringing circuit.

the e_k wave may be assumed to be unaffected by the damping. Thus the period is

$$2\pi \sqrt{(15 \times 10^{-3}) \times 250} = 12.2 \ \mu s$$

(corresponding to a range interval of 2,000 yd), and the amplitude of the first negative peak is

$$14.2 \sqrt{\frac{15 \times 10^{-3}}{250 \times 10^{-6}}} = 110 \text{ volts}$$

The amplitude of the negative pulse in the e_{cn} wave must be great enough to maintain the tube cutoff even at the time of the negative peak of the e_k wave. The plate-to-cathode voltage at the time of this peak is $125 + 110 = 235$ volts, and the corresponding cutoff voltage is -14.5 volts. Thus the amplitude of the input pulse must be at least

$$110 + 14.5 = 124.5 \text{ volts}$$

The oscillation that is started at time t_2 has a positive initial half cycle. Plate current starting in the tube must flow through the capacitor and charge it, because current in the inductance cannot change instantly. The oscillation during the conducting period of V_1 dies away rapidly, because the equivalent circuit for this period is that shown in Fig. 40. Resistor R_{tube} represents the resistance of the conducting tube; and R_{tube},

L, and C effectively form a parallel R-L-C circuit because the steady voltage of the battery has no effect upon the oscillation. The value of R_{tube} may be as high as r_p or as low as $r_p/(\mu + 1)$, depending upon the nature of the grid circuit in Fig. 39A. If the source of e_{cn} has a high internal resistance, e_c is limited almost to zero by grid current and e_{cn} rises and falls with e_k (as in Fig. 39B—positive values of e_{cn} can occur if the voltage across capacitor C_c decreases appreciably while V_1 is cutoff). Thus R_{tube} has a value of about r_p. If the source of e_{cn} has a very low internal resistance, e_{cn} may be held constant after time t_2. The tube then operates as a cathode follower, and the low output resistance $r_p/(\mu + 1)$ damps the oscillations very rapidly.

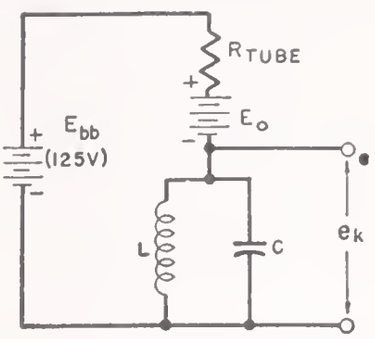


FIG. 40. Equivalent circuit of Fig. 39A when V_1 is conducting.

The R-L-C peaker circuit of Fig. 41A is very similar to the ringing circuit. Differences between Figs. 39A and 41A are as follows: (1) The capacitance of the L-C circuit is usually only the unavoidable inductor and circuit capacitance. Thus the resonant frequency is high. (2) A resistance is connected across the L-C circuit to produce nearly critical

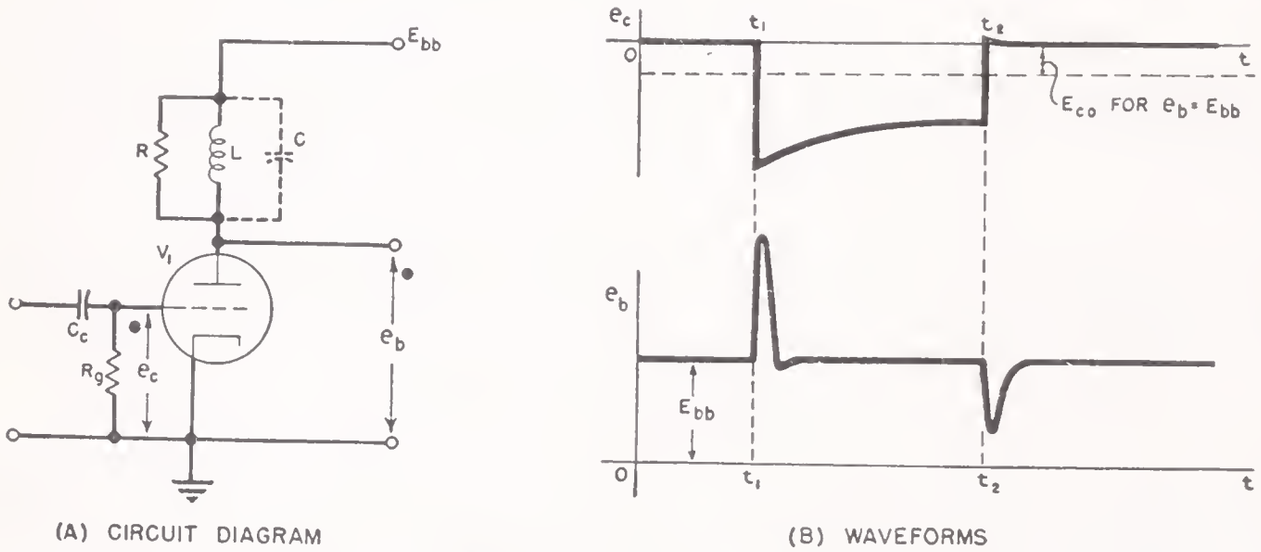


FIG. 41. R-L-C peaker.

damping when the tube is cutoff. Instead of an oscillation, therefore, a single sharp positive peak is produced at time t_1 and a negative peak at time t_2. The negative peak is smaller than the positive one because of the additional damping provided by the conducting tube after time t_2. (3) The inductor is connected to the plate rather than to the cathode of V_1 in order that an output pulse of greater amplitude and steeper leading edge than the input pulse may be obtained. The amplitude of the input pulse must be considerably larger than the cutoff voltage corresponding to $e_b = E_{bb}$, because the voltage necessary for cutoff is increased during the time of the positive pulse in the e_b wave. The applications of this

circuit are similar to those of the *R-C* peaker; use of the inductance makes easier the production of very narrow pulses of large amplitude.

The use of a ringing circuit and an *R-L-C* peaker to form a series of narrow pulses suitable as range markers is illustrated by the circuit of Fig. 42. Tube V_1 is cutoff during the range sweep, so that an oscillation is produced at its cathode during this period. Grid-circuit and plate-circuit clipping in V_2 change the sinusoidal oscillation into a nearly square waveform at the plate of V_2 and grid of V_3. Tube V_3 inverts the square wave and increases the steepness of the edges of the wave. The inductance in the plate circuit of V_4 converts the square wave into the row of pulses shown. The positive pulses are the range markers. Note that the spacing between these (2,000 yd for the circuit values indicated)

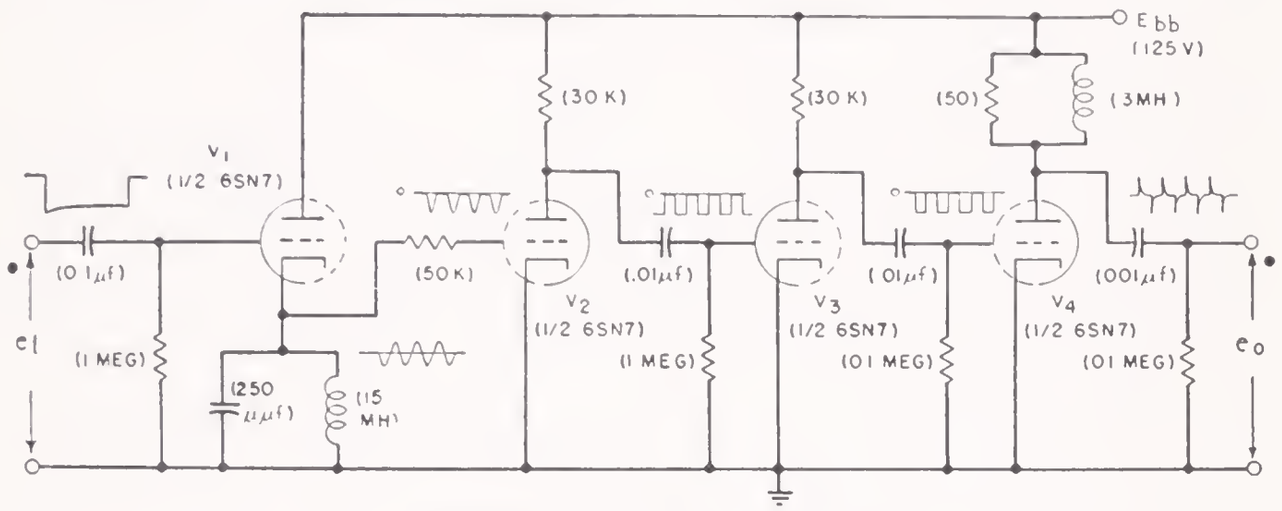


FIG. 42. Range-marker circuit.

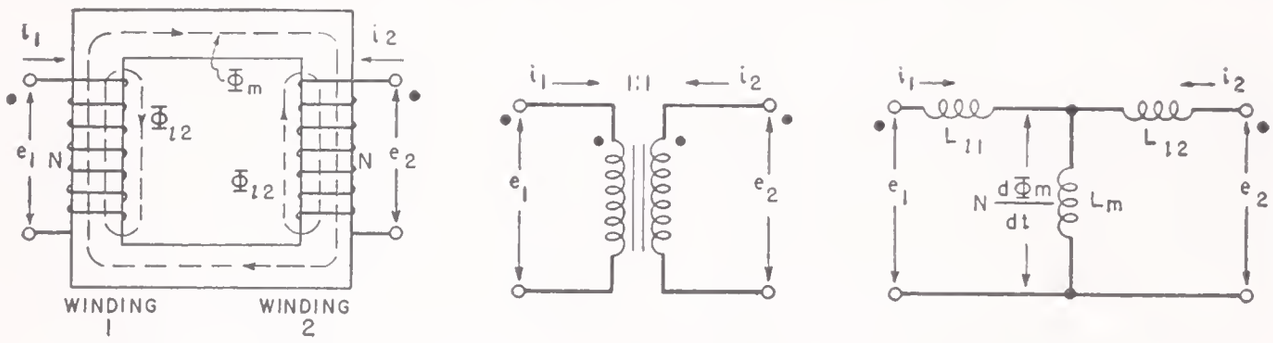
is determined by the inductance and capacitance of the ringing circuit and is little affected by variations of supply voltages or tube characteristics.

If a large number of range markers is to be produced, it may be difficult to obtain a ringing circuit with sufficiently high Q to maintain nearly constant amplitude of oscillation throughout the desired interval. Sometimes the cathode of V_2 is returned through an adjustable resistor to a tap on the ringing-circuit inductance, so that a pulsed Hartley oscillator is formed. The resistor is adjusted so that the feedback exactly compensates for the circuit losses when V_1 is not conducting. Thus oscillations are maintained at constant amplitude during the ringing interval and are rapidly damped when V_1 conducts.

20. Pulse Transformers. Transformers are used in many radar circuits because of their impedance-changing property and because they allow isolation of the primary and secondary circuits. Radar transformers frequently receive rectangular pulses of input voltage; they may be required to reproduce these pulses nearly perfectly or to operate as peakers and reproduce only the pulse edges. To operate effectively with narrow pulses, specially designed *pulse transformers* are required. Pulse

transformers have ferromagnetic cores, are closely coupled, and have relatively few turns in each winding.

Equivalent circuits showing the leakage and magnetizing inductances of a transformer are useful in determining the effect of the transformer on pulses as well as on sine waves. Figure 43*A* shows a diagram[1] of a transformer having the same number of turns N on each winding, Fig. 43*B* shows a circuit-diagram symbol for this transformer, and Fig. 43*C* shows an equivalent circuit in which transformer losses and the isolating property of a transformer are neglected. Because of the currents i_1 and i_2, magnetic flux is set up encircling the windings. The total flux may be divided into three components: (1) the mutual flux Φ_m that threads both windings, (2) the leakage flux Φ_{l1} that links only winding 1, and (3) the leakage flux Φ_{l2} that links only winding 2. Changes of the currents cause



(A) SCHEMATIC DIAGRAM (B) CIRCUIT-DIAGRAM SYMBOL (C) EQUIVALENT CIRCUIT

FIG. 43. One-to-one transformer.

changes of the fluxes, and as a result the voltages e_1 and e_2 are induced in the windings. The dots in Fig. 43*B* indicate the winding direction. The dots are placed so that the voltages induced in the two windings by a given change of mutual flux have the same polarity at the *dot-marked* ends of the windings.

The transformer terminal voltages in Fig. 43*A* each comprise two components, one arising from changes in the mutual flux, the other from changes in one of the leakage fluxes. The component of voltage induced by mutual flux in each winding is $Nd\Phi_m/dt$. The component of e_1 induced by Φ_{l1} is equivalent to the voltage generated in a self-inductance, because Φ_{l1} depends only on i_1. Thus changes of Φ_{l1} generate a voltage $L_{l1}di_1/dt$ in winding 1, where L_{l1} is the leakage inductance of winding 1. In a similar way, the component of e_2 induced by changes of Φ_{l2} is $L_{l2}di_2/dt$, where L_{l2} is the leakage inductance of winding 2.

Reference to Fig. 43*C* shows that, in the equivalent circuit, voltage e_1 is the sum of $L_{l1}di_1/dt$ and $Nd\Phi_m/dt$, and e_2 is the sum of $L_{l2}di_2/dt$ and $Nd\Phi_m/dt$, these components of voltage being the same as in the trans-

[1] Figure 43*A* is a schematic diagram only. Pulse transformers are made with the primary and secondary windings on the same leg of the core in order to reduce leakage flux and leakage inductance.

former itself. In the equivalent circuit $Nd\Phi_m/dt$ is assumed to equal the voltage produced in an inductance L_m (called the magnetizing inductance) by changes of the currents i_1 and i_2. That is, $Nd\Phi_m/dt$ is considered to equal $L_m d(i_1 + i_2)/dt$. To the extent that the magnetization curve of the transformer core is linear, this representation is correct, because Φ_m is to this extent proportional to $i_1 + i_2$ and therefore $Nd\Phi_m/dt$ is proportional to $d(i_1 + i_2)/dt$.

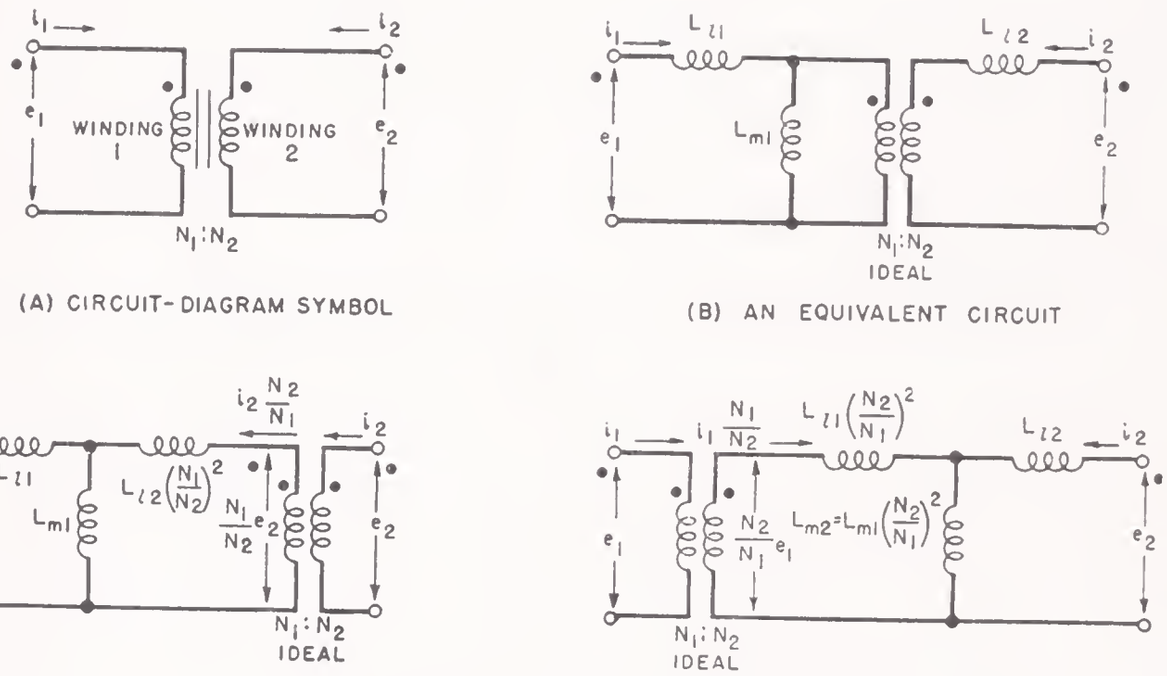
The equivalent circuit shows that $e_2 = e_1$, except for the voltage drops across the leakage inductances. Thus if a transformer is to reproduce pulses or other voltage waveforms nearly perfectly, the leakage inductances should be made as small as possible. In well-designed pulse transformers, the leakage inductances are extremely small, with the result that no appreciable voltage appears across them, except momentarily when very abrupt changes in the input waveform occur. Leakage inductances are like shunt capacitances in that they affect the response to sudden changes and high frequencies.

The equivalent circuit shows also that the current $-i_2$ directed *out of* an output terminal equals the current i_1 directed *into* an input terminal, except for current in the magnetizing inductance. To make i_1 and $-i_2$ nearly equal, L_m should be made as large as possible. Magnetizing inductance does not affect the transformer voltage response directly, but the current in L_m, called magnetizing current, does load the source of input voltage and therefore affects both input and output voltage indirectly. Since current can build up only slowly in a large inductance, the effect of L_m is upon the long-time or low-frequency response of the transformer. Because of the magnetizing inductance a direct current sent through winding 1 produces no permanent voltage across winding 2 and only a very small permanent voltage across winding 1 (the voltage drop across the winding resistance). Pulse transformers intended to operate with pulses of a few microseconds duration or less do not require very large values of L_m. Therefore, few turns are used on their windings in order that the leakage inductances may be small.

Leakage inductances greater than zero and magnetizing inductances less than infinity are unavoidable imperfections in transformers. An imaginary device in which the leakage inductances are zero and the magnetizing inductance is an open circuit is called an *ideal transformer*. A 1:1 turns ratio ideal transformer isolates the circuits of the two windings; that is, a point in the circuit of winding 2 may have any desired voltage with respect to the circuit of winding 1 without effect upon the relation of e_1 to e_2. Otherwise the 1:1 ideal transformer is equivalent to a direct connection between input and output terminals.

If the two windings have different numbers of turns, the principles of operation are not changed but a transformation ratio is introduced

between input and output circuits. If one winding of a transformer has N_1 turns, and if the number of turns on the other winding is changed from N_1 to a greater number N_2, then the voltage induced in the first winding by a given set of changing fluxes is not altered but the voltage induced in the second winding is increased in the ratio N_2/N_1. The increase occurs because voltages induced by changing magnetic flux are proportional to the number of turns linked by the flux. An equivalent circuit for the transformer of Fig. 44*A*, drawn in Fig. 44*B*, is therefore the same as for a 1:1 transformer, except that an $N_1:N_2$ ideal transformer is added to indicate the effect of the turns ratio on voltage e_2.



(A) CIRCUIT-DIAGRAM SYMBOL

(B) AN EQUIVALENT CIRCUIT

(C) ADDITIONAL EQUIVALENT CIRCUITS

FIG. 44. Transformer of any turns ratio with equivalent circuits containing ideal transformer of ratio $N_1:N_2$.

Increasing the number of turns on one winding of a transformer not only increases the voltage that a given flux induces in the winding but also decreases the current in the winding required to set up the given flux, because flux depends upon the product of turns times current. Therefore, in Fig. 44*A*, if N_2 is greater than N_1, the current i_2 in winding 2 is equivalent to a greater current $i_2 \times N_2/N_1$ in a winding of N_1 turns. The ideal transformer in Fig. 44*B* indicates both the current and voltage ratios. Note that in going through an $N_1:N_2$ ideal transformer, current is changed in a ratio that is the reciprocal of the voltage transformation ratio. Thus the product of voltage times current, which is instantaneous power, is not changed and the ideal transformer is lossless.

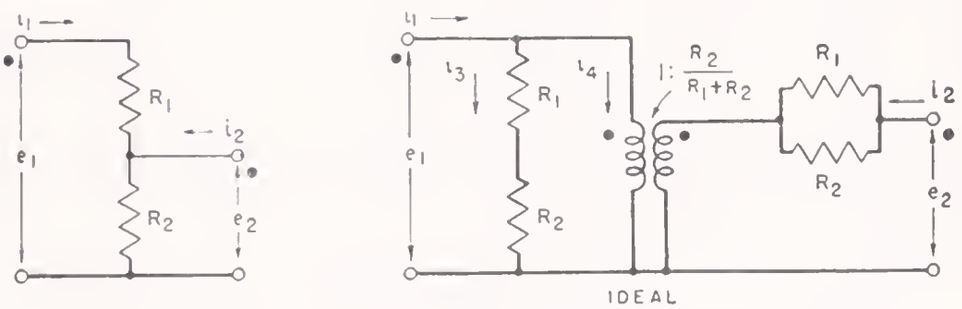
An impedance may be shifted from one side of an ideal transformer to the other, provided that in the shift the impedance is multiplied by the *square* of the transformer turns ratio. For example, if a resistor is con-

nected to the output of a 1:2 step-up transformer, the input impedance of the transformer is (input voltage)/(input current) which is equal to ($\frac{1}{2}$ × resistor voltage)/(2 × resistor current) or $\frac{1}{4}$(resistor voltage)/(resistor current). Thus the resistor connected across the output winding is equivalent to a resistor $(\frac{1}{2})^2$ times as large across the input winding.

As shown in Fig. 44C, the transformer inductances may be shifted to either side of the ideal transformer in the equivalent circuit. In these figures, L_{l1} and L_{m1} stand for the leakage inductance of winding 1 and the transformer magnetizing inductance, respectively, each referred to winding 1. Inductance L_{l2} is the leakage inductance of winding 2. The ratios $(N_1/N_2)^2$ and $(N_2/N_1)^2$ are applied when an inductance is shifted from one side of the ideal transformer to the other. The equivalent circuits in Fig. 44 do not include the losses of the transformer. Power losses result from the resistance of the windings (represented by resistances in series with the leakage inductances in the equivalent circuit) and from hysteresis and eddy currents in the core (accounted for by a resistance in parallel with the magnetizing inductance).

The equivalent circuits in Fig. 44 are by no means the only useful ones for a transformer. Not only may the inductances be transferred from one side of the ideal transformer to the other, but also circuits utilizing ideal transformers of different ratios may be obtained if the inductance values are altered appropriately. A simple artifice of circuit analysis by means of which a great variety of such circuits may be obtained is as follows:

Any voltage divider comprising two similar circuit elements can be replaced by an ideal transformer with shunt and series impedances in the



(A) VOLTAGE DIVIDER (B) IDEAL–TRANSFORMER EQUIVALENT

FIG. 45. Representation of resistance voltage divider in terms of an ideal transformer.

manner illustrated for a resistance voltage divider in Fig. 45. The transformer ratio is the open-circuit ratio of the divider, and the shunt and series resistances are, respectively, the series and shunt combinations of the two resistors of the voltage divider. Thévenin's theorem shows that, for a given applied voltage e_1, both circuits of Fig. 45 supply the same voltage e_2 and current $-i_2$ to whatever load may be present. The two circuits are therefore equivalent, provided the current i_1 is the same for

both. For Fig. 45*A*,

$$e_i = (R_1 + R_2)i_1 + R_2 i_2$$

or

$$i_1 = \frac{e_1}{R_1 + R_2} - i_2 \frac{R_2}{R_1 + R_2}$$

and this relation yields exactly the current i_1 in Fig. 45*B*, $e_1/(R_1 + R_2)$ being the current i_3 and $-i_2R_2/(R_1 + R_2)$ being the current i_4 expressed in terms of i_2 and the transformer ratio.

Voltage dividers comprising two inductors or two capacitors may be represented in an analogous manner. The ideal-transformer equivalent may also be employed when each impedance of the voltage divider is a combination of two or more elements, provided that the two impedances are the same except for a magnitude factor, so that the open-circuit ratio of the transformer is independent of the frequency or waveform of the applied voltage. For example, each impedance of the voltage divider may be a parallel combination of resistance and capacitance, provided that these two *R-C* combinations have the same time constant. The inductance voltage divider is of special interest in the derivation of transformer equivalent circuits.

Four particularly useful circuits obtainable from Fig. 44 by means of the voltage-divider-to-ideal-transformer change are presented in Fig. 46. Figure 46*A* shows the circuit of Fig. 44*B* and the results of changing the divider L_{l1}, L_{m1} in this circuit to an ideal-transformer equivalent. The final circuit is obtained if the parallel L_{l1}, L_{m1} combination is referred to the L_{l2} side of the $N_1:N_2$ transformer and the two ideal transformers combined. The inductance

$$L_1 = L_{m1} + L_{l1}$$

is the *self-inductance* or open-circuit inductance of winding 1. The final ideal-transformer ratio is

$$\frac{L_{m1}}{L_{m1} + L_{l1}} \frac{N_2}{N_1} = \frac{M}{L_1}$$

where $M = L_{m1}(N_2/N_1)$ is the *mutual inductance* of windings 1 and 2. Note from the first equivalent circuit of Fig. 46*A* that for $i_2 = 0$ the voltage across L_{m1} is $L_{m1}di_1/dt$, and thus the winding 2 voltage is

$$e_2 = L_{m1} \frac{di_1}{dt} \frac{N_2}{N_1} = M \frac{di_1}{dt}$$

The inductance L_{sc2} is the short-circuit inductance of winding 2 and is given by

$$L_{sc2} = L_{l2} + \left(\frac{N_2}{N_1}\right)^2 \frac{L_{m1}L_{l1}}{L_{m1} + L_{l1}}$$
$$= L_{l2} + \left(\frac{N_2}{N_1}\right)^2 \frac{L_{m1}(L_{l1} + L_{m1})}{L_{m1} + L_{l1}} - \left(\frac{N_2}{N_1}\right)^2 \frac{L_{m1}L_{m1}}{L_{m1} + L_{l1}}$$
$$= L_{l2} + L_{m2} - \frac{M^2}{L_1} = L_2 - \frac{M^2}{L_1}$$

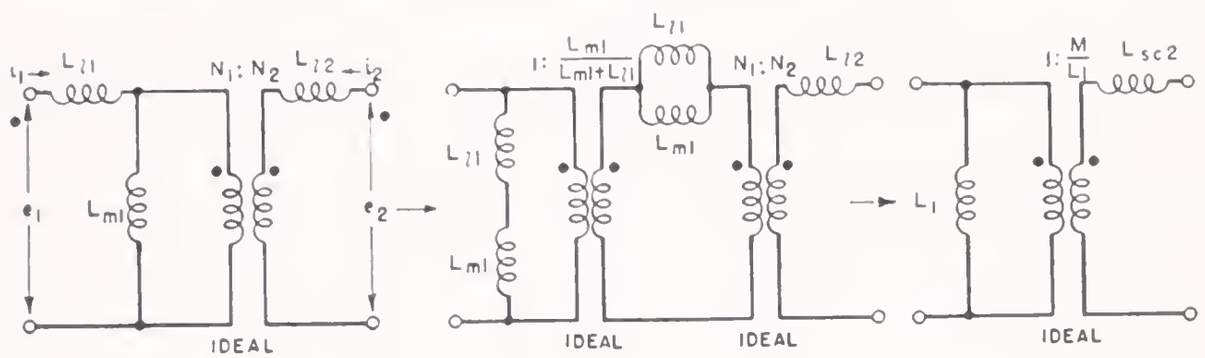
where $L_2 = L_{l2} + L_{m2}$ is the self-inductance of winding 2. Thus

$$L_{sc2} = L_2(1 - k^2) \tag{31}$$

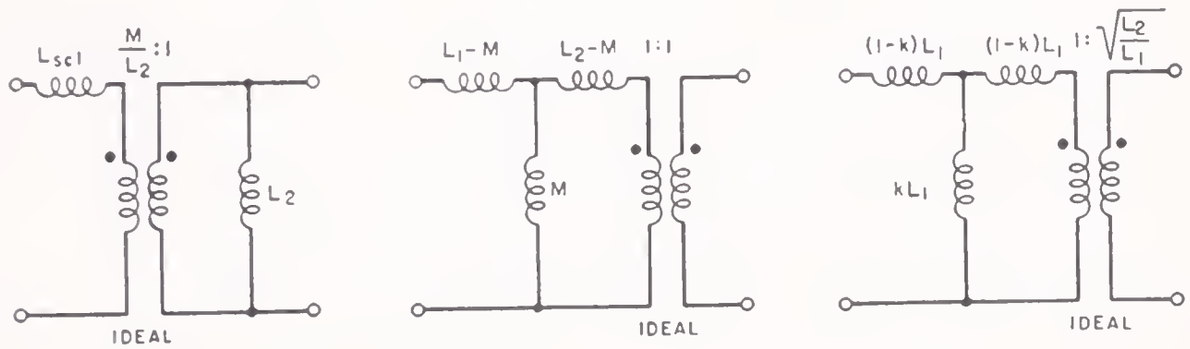
where

$$k = \frac{M}{\sqrt{L_1 L_2}} \tag{32}$$

is the *coefficient of coupling* of the transformer.



(A) $L_1 - L_{sc2}$ CIRCUIT



(B) $L_{sc1} - L_2$ CIRCUIT (C) $L_1 - L_2 - M$ CIRCUIT (D) SYMMETRICAL CIRCUIT

FIG. 46. Additional equivalent circuits for transformer.

The circuit of Fig. 46*B* follows from that of Fig. 46*A* by an interchange of subscripts, and similarly the relations $M = L_{m2}N_1/N_2$ and

$$L_{sc1} = L_1(1 - k^2)$$

are obtained; L_{sc1} is the short-circuit inductance of winding 1. The circuits of Figs. 46*C* and *D* can be reduced to the final circuit in Fig. 46*A* if the voltage dividers $(L_1 - M)$, M and $(1 - k)L_1$, kL_1 are replaced in terms of ideal transformers. The 1 : 1 transformer in Fig. 46*C* is included

merely to represent the isolating property of the transformer. In this circuit either $L_1 - M$ or $L_2 - M$ is usually negative (unless $N_1:N_2$ is approximately unity). In all the other circuits, all inductances are positive because k is always less than unity.

Pulse transformers, being constructed so that leakage inductances are very small relative to magnetizing inductances, have short-circuit inductances that are small fractions of the open-circuit inductances. Therefore, by Eq. 32, k is only slightly less than unity; that is, the transformers are very closely coupled. For pulse transformers the ratios N_2/N_1, M/L_1,

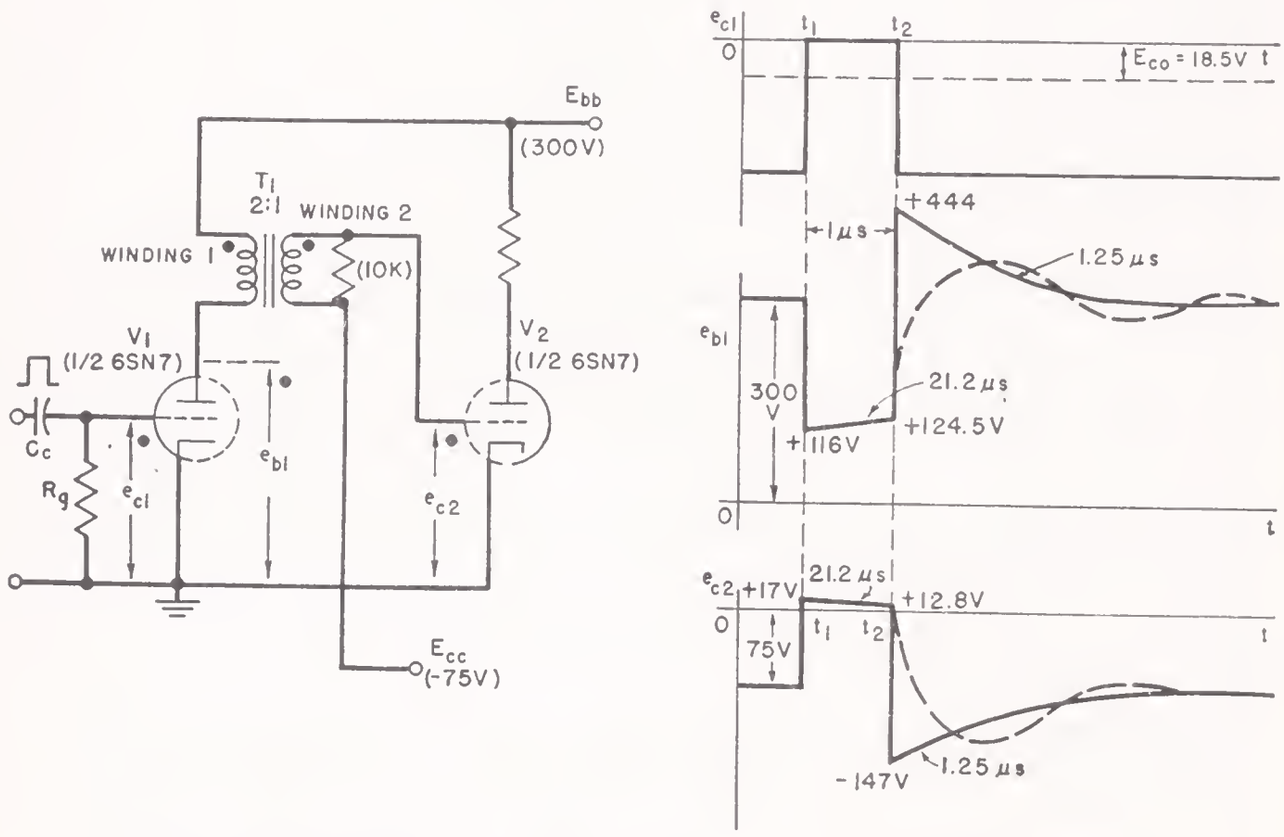


FIG. 47. Pulse transformer coupling two overdriven amplifiers.

L_2/M, and $\sqrt{L_2/L_1}$ are practically indistinguishable. The shunt inductances of Figs. 44*B* and 46*A*, *B*, and *D*, referred to the same winding, may all be considered equal as may the total series inductances. It is therefore immaterial which of these circuits is employed in pulse-transformer analysis; in the following articles the circuit of Fig. 44*B* is adopted.

In later chapters loosely coupled air-core transformers are encountered. For these the parameter values of the various equivalent circuits may differ widely, and it is desirable to employ the circuit that permits the most simple and instructive analysis of the transformer. Sometimes the circuits of Figs. 46*A* and *B* are desirable because they contain two rather than three inductances (see p. 645, Chap. IX); at other times the symmetry of the inductance network of Fig. 46*D* is an advantage (see p. 424, Chap. VI).

Figure 47 illustrates the use of a pulse transformer as a coupling device between two overdriven amplifier stages. Tube V_1 is cutoff, except during 1-μs intervals when a positive pulse is applied to the grid. Negative pulses therefore appear at the plate of V_1, and connections are made to the transformer in such a way that positive pulses are applied to the grid of V_2. Tube V_2 is maintained cutoff except during the pulses by the biasing circuit shown. Use of a transformer makes possible two successive stages in which each of the tubes is cutoff between pulses.

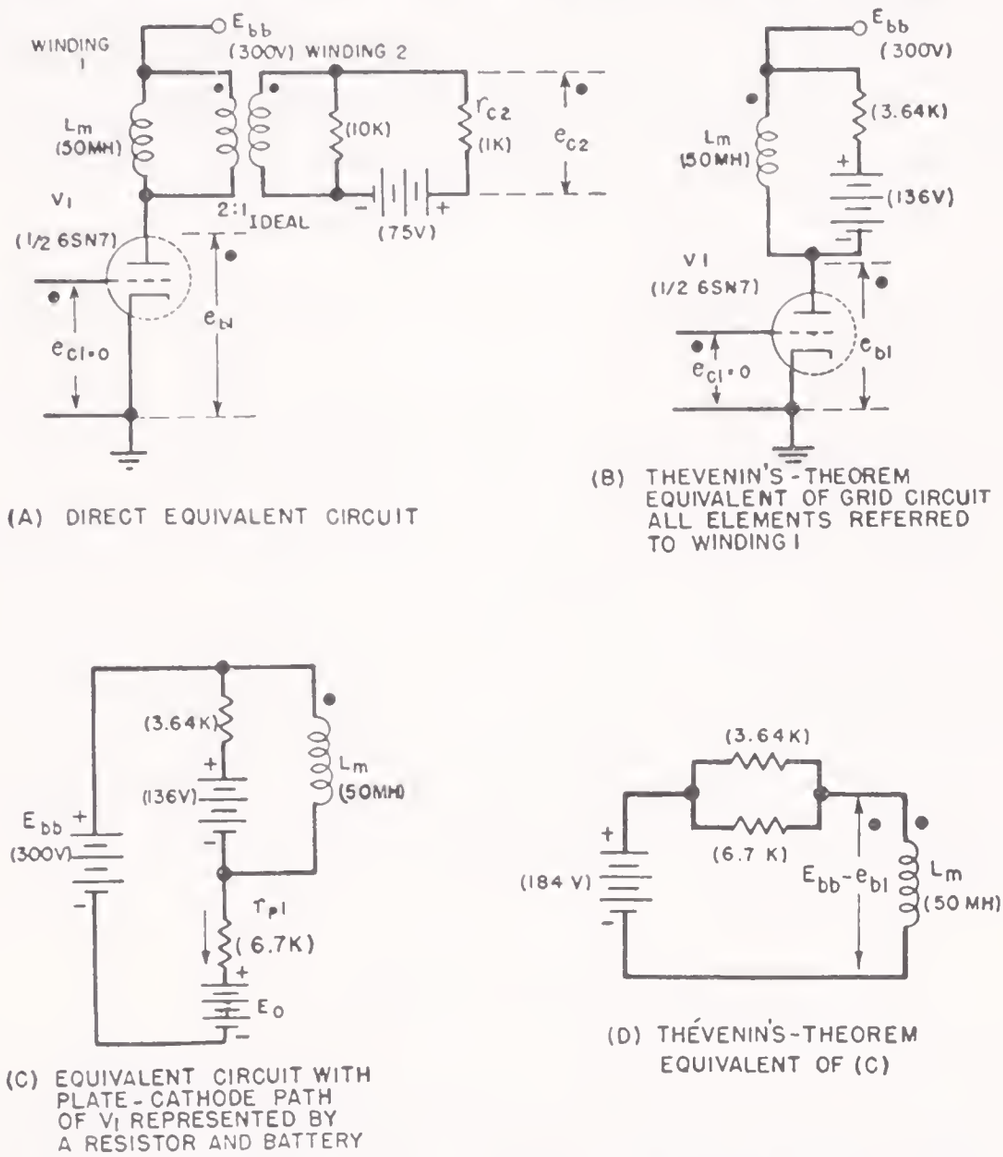


FIG. 48. Equivalent circuits for Fig. 47*A* applicable during pulse.

Assume the circuit constants indicated in Fig. 47*A*, and suppose that the magnetizing inductance of T_1, referred to winding 1, is 50 mh. For simplicity consider first that leakage inductance is completely negligible. Before time t_1, tube V_1 is cutoff; and therefore e_{b1} is 300 volts, no voltage exists across either winding of T_1, and e_{c2} is the bias voltage -75 volts.

The start of current in V_1 at time t_1 causes e_{b1} to drop, and therefore the voltages in each transformer winding are in the direction to make the dot-marked ends positive. The rise of e_{c2} is sufficient to make the grid circuit of tube V_2 conduct; therefore, the equivalent circuit of Fig. 48*A* is applicable between times t_1 and t_2. Figure 48*B* is obtained from Fig.

48*A* by first changing the battery and resistors in the winding 2 circuit into a series combination by means of Thévenin's theorem and then referring the circuit elements to winding 1. The resistance is multiplied by 4, the square of the turns ratio, and the battery voltage by 2, the turns ratio.

When V_1 first begins to conduct, the current in L_m is zero, because an inductor current cannot change instantly. Therefore, the initial value of e_{b1} may be found if L_m in Fig. 48*B* is considered an open circuit, and a load line for the 3.64-K resistor and for a plate supply voltage of $300 - 136 = 164$ volts is drawn. The operating point for $e_c = 0$ on this line is at $e_b = 116$ volts. Thus at time t_1 the e_{b1} wave drops as indicated in Fig. 47*B*, and the transformer winding 1 voltage becomes

$$300 - 116 = 184 \text{ volts}$$

The voltage across winding 2 becomes $\frac{184}{2} = 92$ volts, and e_{c2} rises to $92 - 75 = 17$ volts.

After the initial instant of the pulse, current builds up in the magnetizing inductance L_m, and this added current increases the tube voltage e_{b1} and decreases the transformer voltage. The e_{b1} wave therefore slopes upward and the e_{c2} wave downward during the pulse. To determine the amount of the change, the equivalent circuits of Figs. 48*C* and *D* are used. In Fig. 48*C*, the conducting tube is represented by its equivalent battery-and-resistance combination E_o and $r_p = 6.7$ K. Thévenin's theorem is applied to all of Fig. 48*C* except L_m to obtain Fig. 48*D*. The battery voltage in Fig. 48*D* is the transformer winding 1 voltage when the current in L_m is zero. From Fig. 48*D*, the time constant of the circuit L/R is found to be

$$\frac{50}{(6.7 \times 3.64)/(6.7 + 3.64)} = 21.2 \ \mu\text{s}$$

Therefore the voltage across winding 1 of the transformer decreases from 184 volts to $184\epsilon^{-1/21.2} = 175.5$ volts during the 1-μs pulse, and the plate voltage e_{b1} rises to $300 - 175.5 = 124.5$ volts. The current in L_m at the end of the pulse, obtained from Fig. 48*D*, is

$$\frac{184 - 175.5}{(6.7 \times 3.64)/(6.7 + 3.64)} = 3.6 \text{ ma}$$

At time t_2, the plate current of V_1 is suddenly stopped, and therefore current in winding 1 of the transformer stops equally suddenly. If leakage inductances are neglected, a sudden change of current in one winding of a transformer is not inconsistent with the principle "current in an inductor cannot be changed instantly." However, a simultaneous sudden change of current in the other winding must occur in such a way *that*

no sudden change of the current in L_m occurs. In the 1:1 transformer of Fig. 43*A*, p. 124, if i_1 decreases suddenly by 1 amp, i_2 must simultaneously increase by 1 amp. In this way no sudden change occurs in the sum of the two currents which is the current in L_m or in the flux Φ_m—the magnetic flux cannot change instantly, since the induced voltage $Nd\Phi_m/dt$ cannot be infinite.

Because the current in L_m cannot change from 3.6 ma at the instant when the winding 1 current stops, a current of $3.6 \times 2 = 7.2$ ma must flow in the winding 2 circuit. The direction of this current is opposite to the direction in which grid current can flow. Therefore the equivalent circuit for the period following time t_2 is as drawn in Fig. 49*A*, and this circuit referred to winding 1 is as drawn in Fig. 49*B*. The voltage generated across winding 1 when current in V_1 stops is the voltage produced by

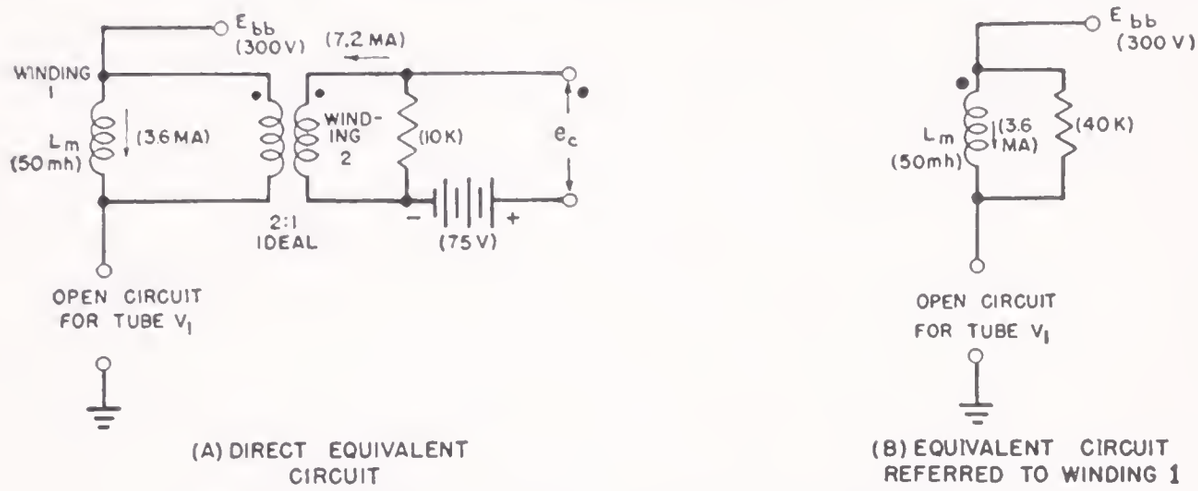


FIG. 49. Equivalent circuits for Fig. 47*A* applicable after pulse.

the 3.6-ma current in the 40-K resistor of Fig. 49*B*; that is, $3.6 \times 40 = 144$ volts. The direction of this voltage is such as to make the dot-marked terminals of the transformer negative. Therefore, e_{b1} rises to

$$300 + 144 = 444 \text{ volts}$$

and e_{c2} drops to $-144\frac{1}{2} - 75 = -147$ volts. Following these initial jumps, e_{b1} decreases exponentially to 300 volts and e_{c2} increases exponentially to -75 volts. The time constant of both exponentials is $\frac{50}{40} = 1.25$ μs.

The solid-line waveforms of Fig. 47*B* show these variations and indicate that the effect of magnetizing inductance is very similar to the effect of the capacitor in an *R-C* coupling circuit. Too small a value of L_m causes the top (or bottom) of the pulse to slope and also causes an overshoot following the pulse. Where flat-topped pulses are desired, pulse transformers having magnetizing inductances of several hundred millihenries are used instead of the 50-mh value of the example. Sometimes transformers with low value of L_m are used with relatively long pulses for the purpose of obtaining a peaking effect.

In the example of Fig. 47, leakage inductances and shunt capacitances have so far been neglected entirely. Leakage inductance and shunt capacitance affect only the high-frequency components of waveforms, that is, only the edges of the pulses in Fig. 47*B*. Because of shunt capacitance, the leading edges of the pulses of e_{b1} and e_{c2} may both be expected to slope a little instead of being perfectly vertical. Leakage inductance causes the edges of the e_{c2} pulse to differ slightly in shape from the edges of the e_{b1} pulse. Often a low-amplitude oscillation of very high frequency appears superimposed on the rise of the leading edges because of oscillations involving the capacitance and leakage inductance. A more important effect occurs at the trailing edge of the pulse because of the oscillations of the magnetizing inductance and shunt capacitance. When tube V_1 in Fig. 47*A* is cut off, the current in the magnetizing inductance flows into the 10-K resistor in parallel with the wiring and transformer capacitances and starts an oscillation. The waveforms shown dotted in Fig. 47*B* result. The 10-K resistor is present to reduce the peak value and increase the damping of this oscillation.

21. Blocking Oscillator. An important and useful circuit for the generation of very narrow pulses is the *blocking oscillator*. Like multivibrators, blocking oscillators may be either free-running or driven. A free-running blocking oscillator determines its own repetition period and generates a series of narrow pulses spaced by this period. It is, therefore, often used as a master oscillator in radar systems. Driven blocking oscillators may be used to convert a broad, rounded trigger pulse into a sharp output pulse, usually of the order of 1-μs duration. Some special blocking oscillators generate very accurately rectangular pulses.[1] Driven blocking oscillators are sometimes used for frequency division.

Similarities exist between the driven blocking oscillator of Fig. 50*A* and the driven multivibrator of Fig. 30, p. 79, the blocking oscillator tube corresponding to the normally off tube of the multivibrator. Before a trigger pulse is applied, the blocking-oscillator tube is maintained cutoff by the bias voltage to which the grid-return resistor is connected. On application of a trigger pulse, a very fast switching action occurs because of regenerative feedback and starts conduction in the tube. In a plate-coupled multivibrator, *R-C* coupling and the normally on tube are used to obtain feedback with the necessary polarity reversal. In the blocking oscillator, a pulse transformer and an *R-C* coupling circuit supply the feedback with phase reversal. The duration of the conducting time for the blocking-oscillator tube depends mainly upon the pulse-transformer magnetizing inductance and the coupling capacitor (C_c in Fig. 50*A*). A second switching action ends the conducting period, as in a multivibrator;

[1] See Art. 13, Chap. III.

the circuit then returns to its initial state and remains quiescent until the next trigger pulse is applied.

Before a trigger pulse is applied, tube V_1 in Fig. 50*A* is cutoff, no voltage exists across the transformer windings, e_b is 300 volts, e_c is -90 volts, and capacitor C_c is charged to 90 volts with the polarity indicated. If a trigger pulse applied through the small capacitor C at time t_1 raises the

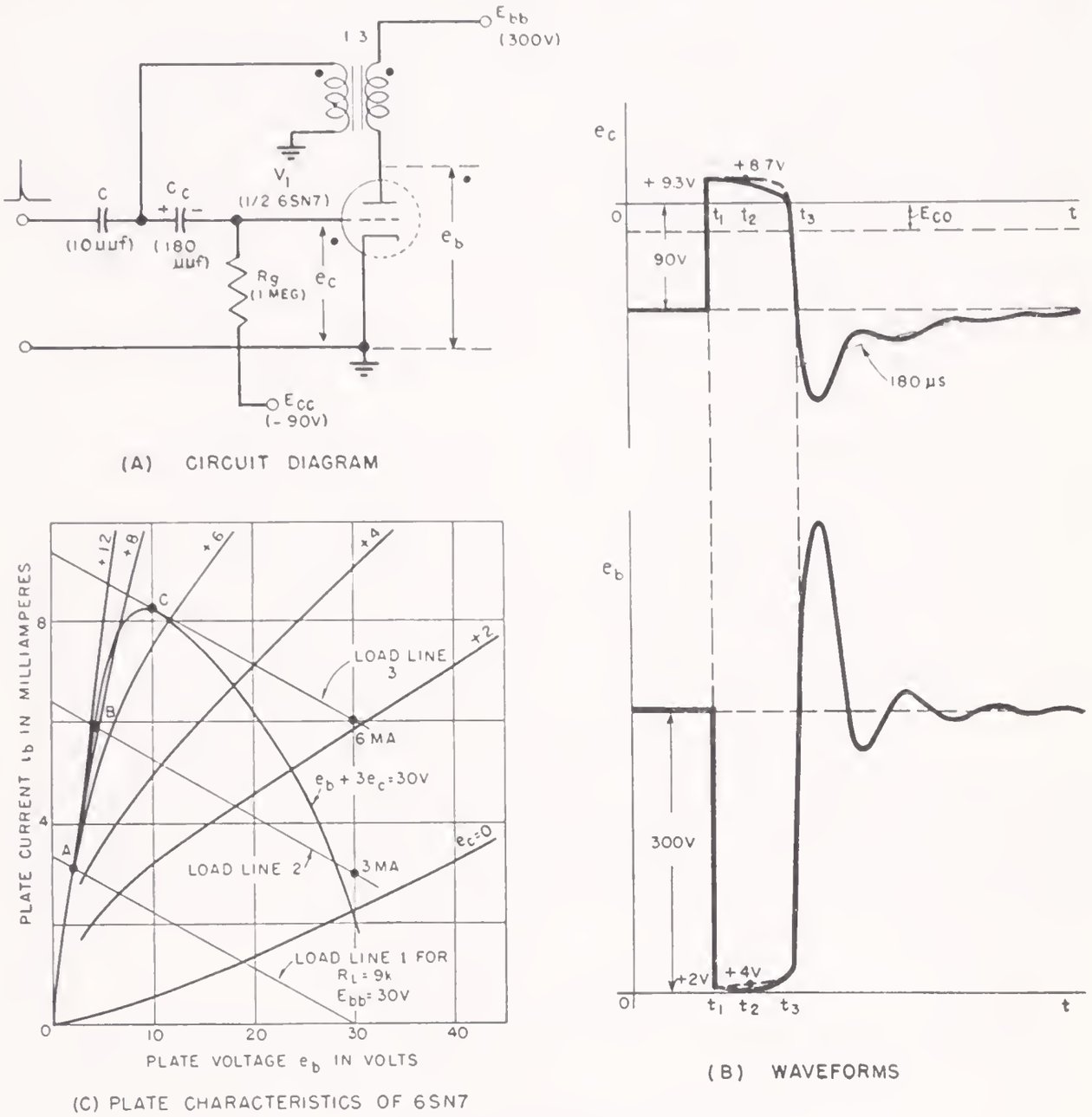


FIG. 50. Blocking oscillator.

grid voltage of V_1 above cutoff, then flow of plate current causes a drop of e_b and, because of the transformer, a further increase of e_c. The process is cumulative, as in a multivibrator, with the result that the tube very quickly begins to conduct a large current and the trigger pulse is no longer needed to sustain the conduction.

The extent of the drop of e_b and rise of e_c when conduction begins can be determined if the equivalent circuit of the transformer and the plate characteristics of the tube are used. In Fig. 51*A* the equivalent circuit

without leakage inductance, shunt capacitance, or trigger capacitance C is drawn, because these elements affect only the shape of the pulse edges. The elements in the equivalent circuit are all referred to the plate winding of the transformer. The grid resistance of the tube is included in the equivalent circuit because at the end of the switching-on process the grid of V_1 is at a large positive voltage. Thévenin's theorem may be used to simplify the circuit of Fig. 51*A*, but no appreciable error is made if the large resistor $9R_g$ is considered an open circuit, as in Fig. 51*B*.

At the first instant after the switching-on process is completed, the current in L_m must be zero and the voltage across C_c must be 90 volts (corresponding to 270 volts across $C_c/9$ in the equivalent circuit). Tube

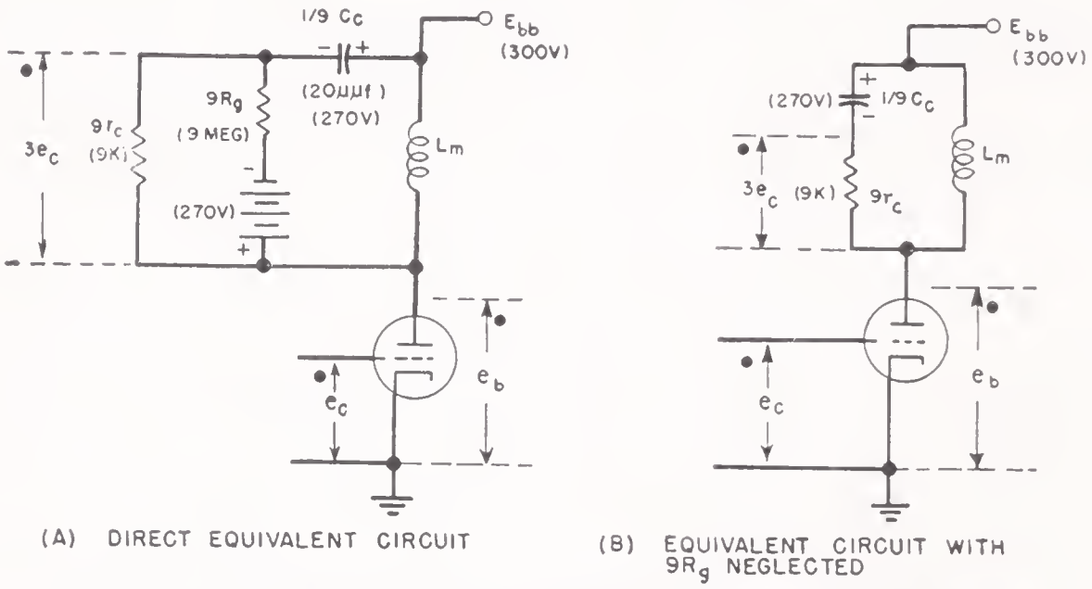


FIG. 51. Equivalent circuits for blocking oscillator during pulse.

V_1, therefore, operates on load line 1 in Fig. 50*C*, which is drawn for a voltage of 300 − 270 = 30 volts and a resistance of 9 K. The point of operation on this line depends upon the value of e_c. It is the point at which the relation

$$e_b + 3e_c = 300 - 270 = 30 \text{ volts} \tag{33}$$

is satisfied. This relation is obtained by adding voltages in the circuit of Fig. 51*B*. The point of operation may be found by a trial-and-error calculation or by plotting the curve $e_b + 3e_c = 30$ volts on the plate characteristics, as in Fig. 50*C*. Point *A*, the intersection of the curve and load line, is the end point of the switching-on process; at this point $e_b = 2$ volts and $e_c = (30 - 2)/3 = 9.3$ volts.[1]

Following the time t_1, the point of operation moves away from point *A* for two reasons: (1) At point *A* current is supplied from the transformer through C_c to the grid of V_1. This current charges C_c, and therefore e_c

[1] A large step-down ratio for the pulse transformer in Fig. 50*A* and a bias voltage of unusually large magnitude were chosen in order to obtain voltage values convenient for calculation. In actual circuits positive grid voltages of 50 to 100 volts are usual.

tends to decrease. (2) At point A a voltage exists across the transformer magnetizing inductance. This voltage causes the magnetizing current to increase, and therefore the plate current of the tube tends to increase. The changes of voltage and current produced by both of these effects are slow compared with the changes that occur during the switching-on process.

For simplicity, assume first that C_c is so large that the change of capacitor voltage can be neglected, and consider only the effects of the increase of magnetizing current. Suppose that at time t_2 current in L_m has increased to 3 ma. Then for any value of plate voltage maintained by the tube, current in the transformer plate winding must be 3 ma greater than the value specified by the original load line; that is, the tube operates on load line 2 in Fig. 50C, which line is parallel to the original load line but 3 ma higher. Equation 33 still applies, because the voltage across $C_c/9$ is unchanged. The new operating point is therefore intersection B. At this point $e_b = 4$ volts and $e_c = (30 - 4)/3 = 8.7$ volts. These values are indicated on the dotted curve in Fig. 50B.

As current continues to increase in L_m, the operating point moves up the curve $e_b + 3e_c = 30$ volts until point C on load line 3 (corresponding to a current of 6 ma in L_m) is reached. This point marks the end of the pulse because a further increase of current in L_m cannot be balanced by a greater plate current obtained by an increase of e_b. Instead an increase of e_b results in a lower transformer voltage, a lower value of e_c, and a lower value of i_b. The decrease of i_b causes a further increase of e_b, and the regenerative switching-off process is started.

Because of the shape of the curve $e_b + 3e_c = 30$ volts, the plate voltage and grid voltage are nearly constant during most of the pulse (see dotted curves in Fig. 50B), and the plate current increases as the magnetizing current increases. The constancy of the plate voltage makes simple the relation between pulse duration and L_m. The average voltage across L_m during the pulse is approximately $300 - 4 = 296$ volts. Therefore if $L_m = 100$ mh, the average rate of change of magnetizing current is

$$\frac{di}{dt} = \frac{e}{L} = \frac{296}{100} = 2.96 \text{ ma per } \mu\text{s}$$

Because the magnetizing current increases from 0 to 6 ma during the pulse, the pulse duration would be $6/2.96 = 2.03$ μs if capacitor C_c were so large as to have no appreciable effect upon pulse duration.

The values of C_c used in actual circuits are almost always small enough to affect the pulse duration. If the voltage across C_c changes during the pulse, the operating point on the plate characteristics is affected in two ways: (1) As the capacitor $C_c/9$ in Fig. 51B charges, the sum of e_b and $3e_c$ decreases. Thus in Fig. 50C the curve, $e_b + 3e_c =$ a constant, moves

toward the origin. (2) As $C_c/9$ charges, the effective plate supply voltage used in drawing the load line is decreased. Thus the load line in Fig. 50*C* moves parallel to itself and to the left.

To the horizontal motion of the load line caused by the change of capacitor voltage must be added the upward motion caused by the growth of current in L_m. The combined effects of capacitor-voltage changes and magnetizing current may be summarized as follows:

1. The grid voltage decreases during the pulse. The decrease is slight if C_c is large but may be considerable if C_c is small enough to charge appreciably during the pulse.

2. The plate voltage decreases at the beginning of the pulse and increases a little near the end. The decrease is slight if C_c is large but may extend for most of the pulse duration if C_c is small.

3. The pulse duration is less than that calculated in terms of magnetizing inductance alone. The smaller the value of C_c, the greater its effect in reducing the pulse duration.

4. At the end of the pulse, a magnetizing current flows in the transformer, and an excess voltage is stored on C_c. The smaller the value of C_c, the smaller the current in the inductance, and the larger the voltage across C_c.

The solid curve in Fig. 50*B* is drawn for an intermediate value of C_c.

The waveform of e_b after the cutoff time t_3 is the same as the corresponding waveform in an overdriven transformer-coupled amplifier. The magnetizing current causes an overshoot of e_c and e_b, and because of shunt capacitance, an oscillation may often result. Often a resistor is connected across the transformer to damp this oscillation. The overshoot or oscillation in the grid-voltage waveform is superimposed upon an exponential rise toward the bias voltage, -90 volts. This exponential change results as the charge added to C_c during the pulse flows through the high resistance of R_g. The time constant of this exponential is R_gC_c, or 180 μs for the circuit of Fig. 50*A*.

Because of the charge stored on C_c during the pulse, a bias voltage is not necessary in a driven blocking oscillator. If the time constant R_gC_c is large enough so that e_c remains below cutoff during the periods between trigger pulses, the oscillator operates in synchronism with the driving voltage. Because of the transformer, negative trigger pulses applied to the plate of the blocking-oscillator tube may be used instead of positive pulses applied to the grid in Fig. 50*A*.

If R_g is returned to the cathode of V_1, or to a positive voltage, and if no trigger pulses are applied, a free-running blocking oscillator is obtained. A pulse is produced each time the grid voltage rises to the cutoff value. The repetition period is difficult to calculate because the voltage to which C_c is charged depends upon the relation between C_c and L_m.

22. Traveling Waves on Artificial Transmission Lines. Networks of inductance and capacitance that behave like sections of transmission lines are often used in radar to form rectangular pulses, to delay pulses, or to modify the shapes of pulses in special ways. These networks are used instead of actual transmission lines because the length of real line required is too long to be practical—about 1,000 ft of line would be needed to delay a pulse by 1 μs. Also, in many applications, the characteristic resistance of real lines is too low for convenient use with tubes.

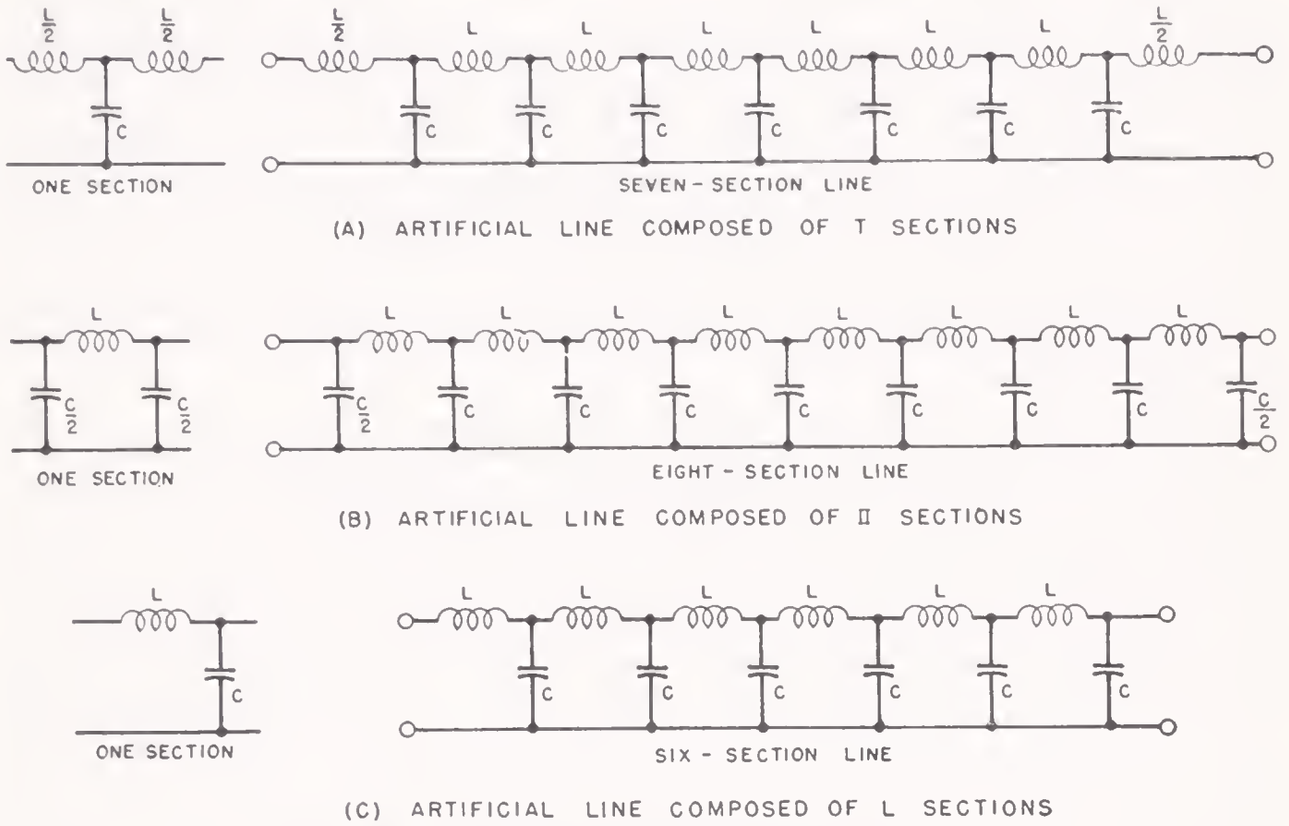


FIG. 52. Examples of artificial transmission lines.

The artificial lines of Fig. 52 are examples of one type of network often used to simulate transmission lines. An artificial line is a chain of T, Π, or L sections, each section containing series inductance and shunt capacitance.

An artificial line having many sections with very small inductance and capacitance in each evidently resembles a real line which has inductance and capacitance distributed along its length. If a sinusoidal voltage wave is applied to an artificial line, and if the frequency of the applied voltage is low so that the phase of the wave changes only a few degrees from one section to the next, then the operation of the line is essentially the same as if the inductance and capacitance were uniformly distributed along the line. An analogy exists between the electrical network and a vibrating string loaded with uniformly spaced weights—if the wavelength of the vibration of the string spans many weights, the resultant motion is as if the string had a large and evenly distributed mass.

If the frequency of the voltage applied to the network is made too high, so that the wavelength along the network is short compared with one section, the similarity to a smooth line disappears. The critical frequency below which the operation of an artificial line approximates that of a real line is called the *cutoff frequency* f_{co} and is given by[1]

$$f_{co} = \frac{1}{\pi \sqrt{LC}} \tag{34}$$

A pulse applied to a lossless real line travels along the line at a definite velocity and with no change of shape. Essentially the same thing happens when a pulse is applied to an artificial line, provided that all the important Fourier-series components of the pulse lie below the cutoff frequency of the line. The artificial line transmits frequencies between 0 and f_{co} and therefore has an effect on pulses much like that of a receiver video amplifier having a bandwidth $\Delta f = f_{co}$. The amount of distortion of the pulse depends upon the relation of f_{co} to $1/T$, where T is the pulse duration.[2]

Figure 53 is drawn to show the mechanism by which a steep wavefront travels along a real transmission line. The switch S is closed to connect the battery of voltage E to the input terminals of the line. The voltage E does not appear instantly at all points along the line. Instead, a wave of voltage (shown in the figure) progresses along the line; the farther a given point is from the battery, the later the time at which the line voltage at that point jumps from 0 to E. A current wave travels away from the battery exactly in step with the voltage wave; at all points up to the moving wavefront, a current I flows away from the battery in the top conductor and a return current I flows in the bottom conductor. Plus and minus charges appear on the conductors at each point where voltage exists between conductors, and magnetic flux lines encircle the conductors wherever current flows. The charges and flux lines are indicated in Fig. 53.

The reason that the line, unlike a capacitor of small dimensions, cannot charge all at once is that the series inductance of the line (associated with the magnetic flux) as well as the shunt capacitance (associated with the electric charge) is important. The voltage wave can progress only as fast as the line current can carry charge to the wavefront to produce the change in voltage. The current wave can progress only as fast as the voltage that appears across a short section of each conductor at the wave-

[1] The artificial line is a multisection low-pass constant-k filter. The cutoff angular frequency is therefore $\omega_{co} = 2/\sqrt{LC}$, which is equivalent to Eq. 34. See, for example, E. A. GUILLEMIN, *Communication Networks*, Vol. II, 279–377, especially p. 317 (John Wiley & Sons, Inc., New York, 1935).

[2] See Arts. 3 and 7, Chap. VI.

front can start current in the corresponding section of line inductance. Therefore the voltage and current waves must move together along the line.

Let τ be the time required for the waves to travel a unit of length along the line, and let c and l be the capacitance and inductance per unit length of the line. Then $I\tau$ is the charge that flows from the battery into the

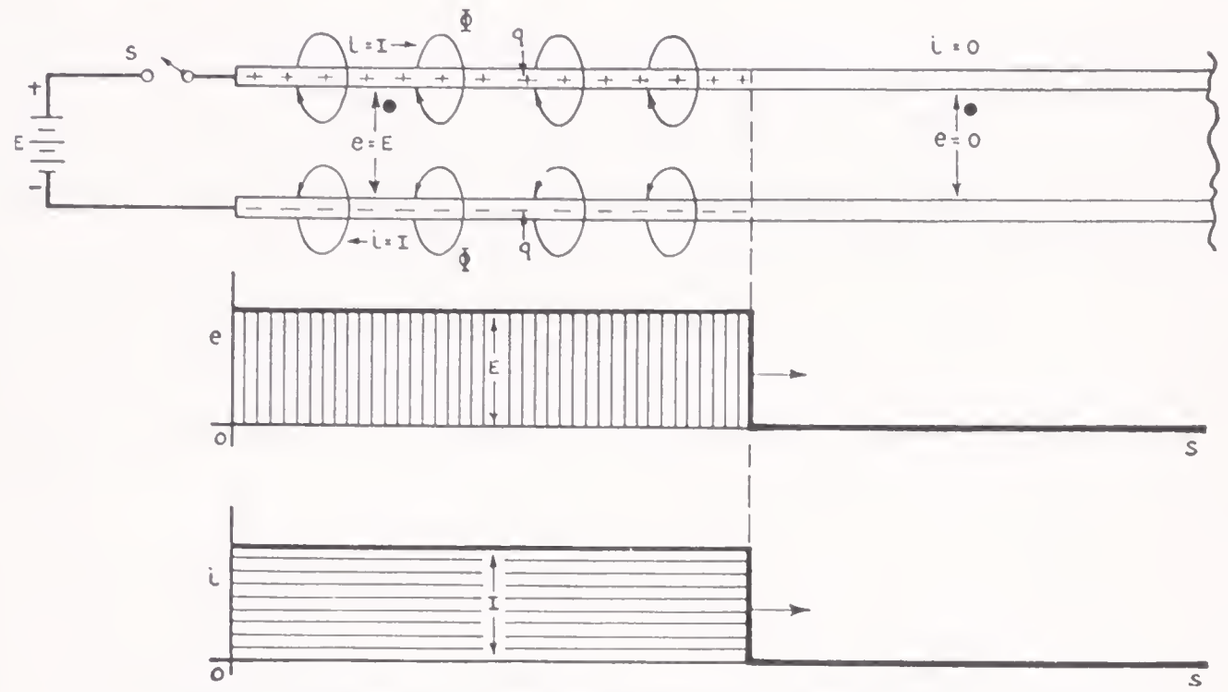


FIG. 53. Propagation of a steep wavefront on a transmission line.

line in the time interval τ, and cE is the charge accumulated on the line at the moving wavefront during the same time interval. Therefore

$$I\tau = cE \tag{35}$$

During the same time interval τ, the flux encircling the conductors increases by an amount lI.* The voltage E at the wavefront must be generated by the changing flux, and therefore $E = d(N\Phi)/dt$ is the rate at which flux linkages are being produced at the wavefront, or $E\tau$ is the increase of flux linkages in the interval τ. Therefore

$$E\tau = lI \tag{36}$$

If Eq. 35 is multiplied by Eq. 36 and the common factors EI canceled, the value of τ is obtained:

$$\tau^2 = lc$$

and

$$\tau = \sqrt{lc} \tag{37}$$

* Inductance can be defined as flux linkages per unit current $(N\Phi/I)$; therefore lI is the number of flux linkages associated with the inductance l of a unit length when the current I is flowing in it.

That is, the wave progresses one unit of length in a time interval $\sqrt{lc}$, or the speed at which the wave travels is $1/\sqrt{lc}$. If Eq. 36 is divided by Eq. 35 and the result multiplied by E/I, the ratio of E to I is obtained:

$$\left(\frac{E}{I}\right)^2 = \frac{l}{c}$$

and

$$\frac{E}{I} = R_c = \sqrt{\frac{l}{c}} \tag{38}$$

The ratio E/I is called R_c, the *characteristic resistance* of the line. The current drawn by the line from the battery is the same as if a resistance R_c were connected to the battery terminals.[1]

If in Fig. 53 an initial voltage had existed between the line conductors or an initial current had been flowing in the conductors, no essential difference in the performance of the line would have resulted. The waves would still travel at the same speed, and the ratio of *change* in voltage to *change* in current would still be R_c, because the transmission line is a linear device. Note and remember:

Any change that occurs on a transmission line takes the form of a voltage wave and a current wave traveling together in one direction or the other or of two such pairs of waves traveling in opposite directions simultaneously. The speed of each wave is $1/\sqrt{lc}$, *and the ratio of each voltage wave to the corresponding current wave is* $\sqrt{l/c}$.

Waves having frequency components below the cutoff frequency travel along artificial lines in the same manner as along real lines. The time of travel and characteristic resistance in terms of the artificial-line parameters are

$$\tau = \sqrt{LC} \tag{39}$$

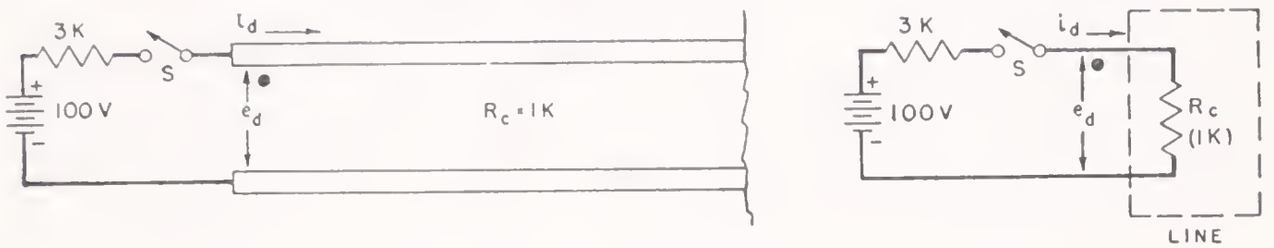
$$R_c = \sqrt{\frac{L}{C}} \tag{40}$$

where L and C are the inductance and capacitance per section of the network (see Fig. 52), τ is the travel time per section, and R_c is the characteristic resistance. These values are accurate only for components of the wave having frequencies very low compared with the cutoff frequency. Components of frequencies slightly lower than f_{co} travel along the line, but with different values of τ and R_c. Components of frequency greater than f_{co} are not propagated along the line. The variations of τ and R_c with frequency cause distortion of pulses in addition to that produced by elimination of frequencies above f_{co}.

[1] Compare the results of this paragraph with the discussion of traveling sine waves on transmission lines in Art. 2, Chap. VII. Waves of any shape travel with the same velocity and same ratio of voltage to current on lossless transmission lines.

Networks entirely different from the artificial lines of Fig. 52 are frequently used to simulate transmission lines. Some examples are given in Art. 4 of Chap. III. Every network that simulates a transmission line has a characteristic resistance and a travel time and operates in a manner that may be interpreted in terms of the transmission line it simulates. Networks different from the artificial-line networks in Fig. 52 are used in order to reduce the distortion of pulses without increasing the number of elements.

The major differences between the effects of simulative networks and real lines upon pulses occur at the steep edges of the pulses, because these give rise to the highest frequency Fourier-series components. If a voltage source is suddenly connected to a real line, the line behaves as a resistor R_c in the first instant. If a simulative network replaces the real line, the network behaves momentarily as either a capacitor or an inductor,



(A) CIRCUIT DIAGRAM (B) EQUIVALENT CIRCUIT

FIG. 54. Line with zero initial current and voltage.

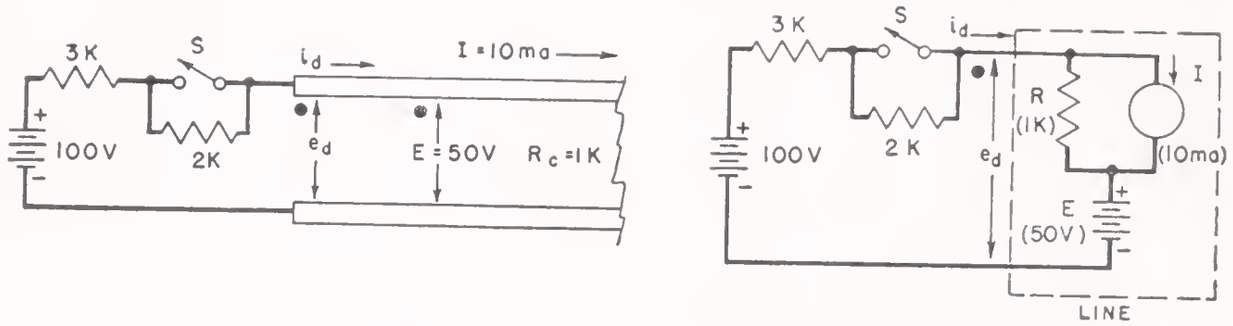
depending upon which element is connected to the input terminal of the network. Very shortly after the first instant, however, the network approximates the real line and behaves as a resistor R_c. Uniform artificial lines cause damped oscillations of very high frequency to appear after each pulse edge; some other types of simulative networks avoid these oscillations.

23. Switching Operations in Transmission-line Circuits. In Art. 22 and Fig. 53, a voltage E is connected directly to a transmission line when a switch is closed. The voltage of the wave propagated along the line is E, and the current of the wave is E divided by the characteristic resistance of the line. When the input circuit is more complicated than in Fig. 53, or when an initial voltage and current are present on the line, a calculation is required to determine the magnitudes of the voltage and current waves that result from a switching operation. In this calculation, the line is assumed to be a real transmission line. Approximately the same results are obtained with artificial lines.

Consider first the circuit of Fig. 54*A* in which a 100-volt battery is connected to a 1-K line through a 3-K resistor when switch S is closed. Assume that the line is uncharged and carries no current previous to the instant of closing the switch. Closing S causes sudden changes of e_d and

i_d (the voltage and current at the line terminals) from zero, and the ratio of the change of e_d to the change of i_d is $R_c = 1$ K. Thus the line is equivalent to a resistor R_c in its effect upon the input circuit, and an equivalent circuit for the switching operation is the one in Fig. 54*B*. From the equivalent circuit, the values of e_d and i_d may be found: i_d is $100/(3 + 1) = 25$ ma, and e_d is $25 \times 1 = 25$ volts. Therefore, when S is closed, the voltage at the line terminals jumps from 0 to 25 volts, and a 25-volt wavefront progresses along the line. Accompanying this voltage wave is a 25-ma current wave.

In the circuit of Fig. 55*A*, a condition is assumed to have been reached with S open in which the voltage $E = 50$ volts appears between conductors at each point of a section of the line near the input terminals, and the current $I = 10$ ma flows in the conductors throughout the section. Note that these initial values are consistent with the battery voltage and



(A) CIRCUIT DIAGRAM (B) EQUIVALENT CIRCUIT

FIG. 55. Line with initial current and voltage.

the resistance magnitudes indicated in the figure. The effect of closing S, taking into account the energy stored in the line due to line-inductance current and line-capacitance voltage, is to be found.

Closing S causes a *change* of e_d which, as before, is R_c times the *change* of i_d. However, since e_d changes from 50 volts and i_d from 10 ma, the line must be represented by the equivalent circuit of Fig. 55*B* rather than by R_c alone. This equivalent circuit is correct because (1) it yields the values of $e_d = 50$ volts and $i_d = 10$ ma before S is closed and (2) it indicates that closure of S causes e_d to change by R_c times the change of i_d. From this circuit the values of e_d and i_d after S is closed may be calculated. One simple procedure is to apply Thévenin's theorem to the parallel combination of R_c and the 10-ma current source. This combination is thereby replaced by R_c in series with a voltage source $IR_c = 10$ volts, the lower terminal of the voltage source being positive. In the resulting over-all series circuit, the net voltage is 60 volts, and the total resistance (with S closed) is 4 K. Thus, i_d is $\frac{60}{4} = 15$ ma, and e_d is

$$100 - (3 \times 15) = 55 \text{ volts}$$

Consequently, the wavefront that is started along the line by closing S

causes a voltage change of 55 − 50 = 5 volts and a current change of 15 − 10 = 5 ma.

A line having initial voltage but no initial current is illustrated in Fig. 56*A*. Assume that, before *S* is closed, the entire line is charged to a voltage $E = 100$ volts. When *S* is closed, current flows into the 3-K resistor as the line capacitance discharges. In the transmission-line equivalent circuit forming part of Fig. 56*B*, the battery representing the initial line voltage is shown, but the current source is omitted because there is no initial current. Calculations on the equivalent circuit show that when *S* is closed i_d becomes $-100/(3 + 1) = -25$ ma and e_d drops from 100 to $25 \times 3 = 75$ volts. The negative value of i_d indicates that current flows in a direction opposite to the i_d arrow of Fig. 56*A*, just as it would if the line were a capacitor initially charged to the same voltage.

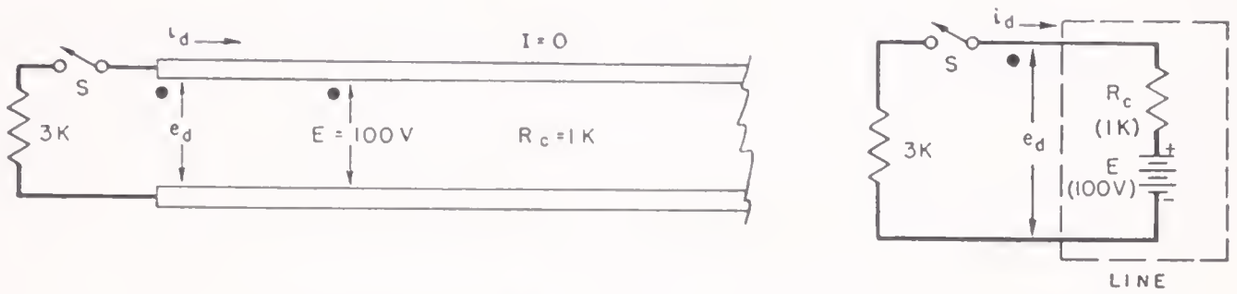


FIG. 56. Discharge of line with initial voltage.

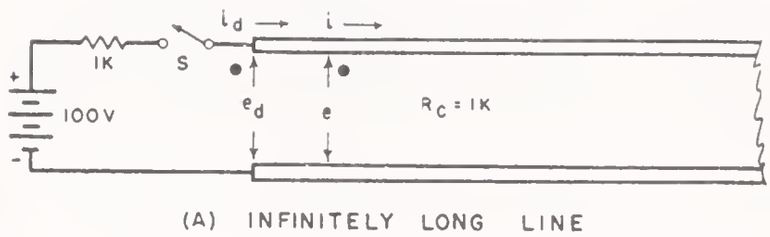
The effect of closing the switch may be interpreted as a negative wavefront of 25-volt and 25-ma amplitude that travels away from the input terminals.

24. Reflections at Line Terminals. The processes thus far considered are the start of a wave from the input terminals of a line and the propagation of a wave along a line of indefinite length. The effects that occur when the wave reaches the output terminals of the line also are of interest.

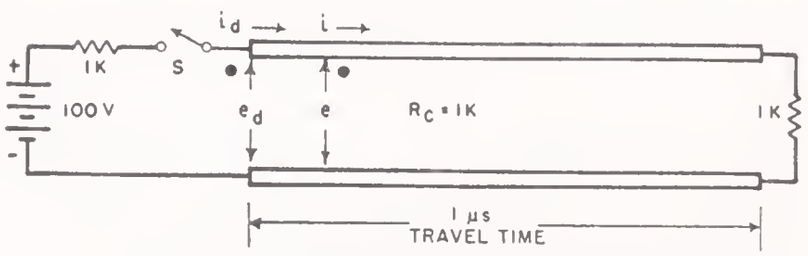
Consider the circuits of Fig. 57. In Fig. 57*A*, a 100-volt battery is connected through a 1-K resistor to an infinitely long line having a characteristic resistance of 1 K. The circuit of Fig. 57*B* is the same except that a line section having a 1-μs travel time is used, and the section is terminated in a 1-K resistor. In either circuit, a 50-volt, 50-ma wave travels to the right from the input terminals after the switch is closed. The operation of the line section of Fig. 57*B* must be identical with the operation of the first 1-μs section of the infinite line, because the remainder of the infinite line is exactly equivalent to a 1-K resistor in its effect upon the first section. Note that in Fig. 57*B* steady-state operation is obtained as soon as the wave reaches the terminating resistor. In the steady state, 50 volts appear between the line conductors and 50 ma flow through the

circuit, as might be expected when two 1-K resistors are connected in series across a 100-volt battery.

The circuit of Fig. 58*A* differs from the one in Fig. 57*B* in that the far end of the transmission line is open-circuited. When switch *S* is first closed, the line appears as a resistance R_c connected to the battery circuit, because the terminal conditions can have no effect before the wavefront reaches the end of the line. Thus a 50-volt, 50-ma wave starts from the source as before. At the end of the line, the *current must be zero at all times* because the terminals are open-circuited. To maintain this zero-current condition, at the instant the 50-ma current wave reaches the open



(A) INFINITELY LONG LINE



(B) LINE TERMINATED IN ITS CHARACTERISTIC RESISTANCE

FIG. 57. Comparison of infinitely long line to line terminated in its characteristic resistance.

end a second 50-ma wave must start from the open end and progress toward the battery. Associated with this *reflected* current wave must be a reflected voltage wave. Because the characteristic resistance must relate voltage and current waves, the voltage-wave amplitude is 50 volts and the total line voltage rises to 100 volts. These traveling waves are indicated in Fig. 58*A*, and the corresponding time variations of e_d and i_d are shown in Fig. 58*B*.

The *incident* wave in Fig. 58*A*, in traveling from the battery to the open end of the line, charges the line capacitance from 0 to 50 volts because of the 50-ma current supplied by the battery. When the wave reaches the open end, no capacitance remains to be charged. The 50-ma current is maintained by the inductance of the line and charges the capacitance near the open end to a higher voltage. Thus a voltage is produced across a section of line inductance near the open end in the direction to reduce the line current. The increase of voltage and decrease of current near the open end of the line corresponds to a reflected wave

starting toward the battery. As the reflected wave progresses, the higher voltage on the open-end side of the wavefront causes the line current to decrease from 50 ma to zero at the wavefront and the 50-ma reflected current wave charges the line capacitance at the wavefront and causes the increase of voltage. When the reflected wave reaches the battery, the entire line is charged to 100 volts and the current is zero everywhere on the line. Therefore the circuit comprises a 100-volt battery connected to a line having a static charge of 100 volts, and no further change occurs.

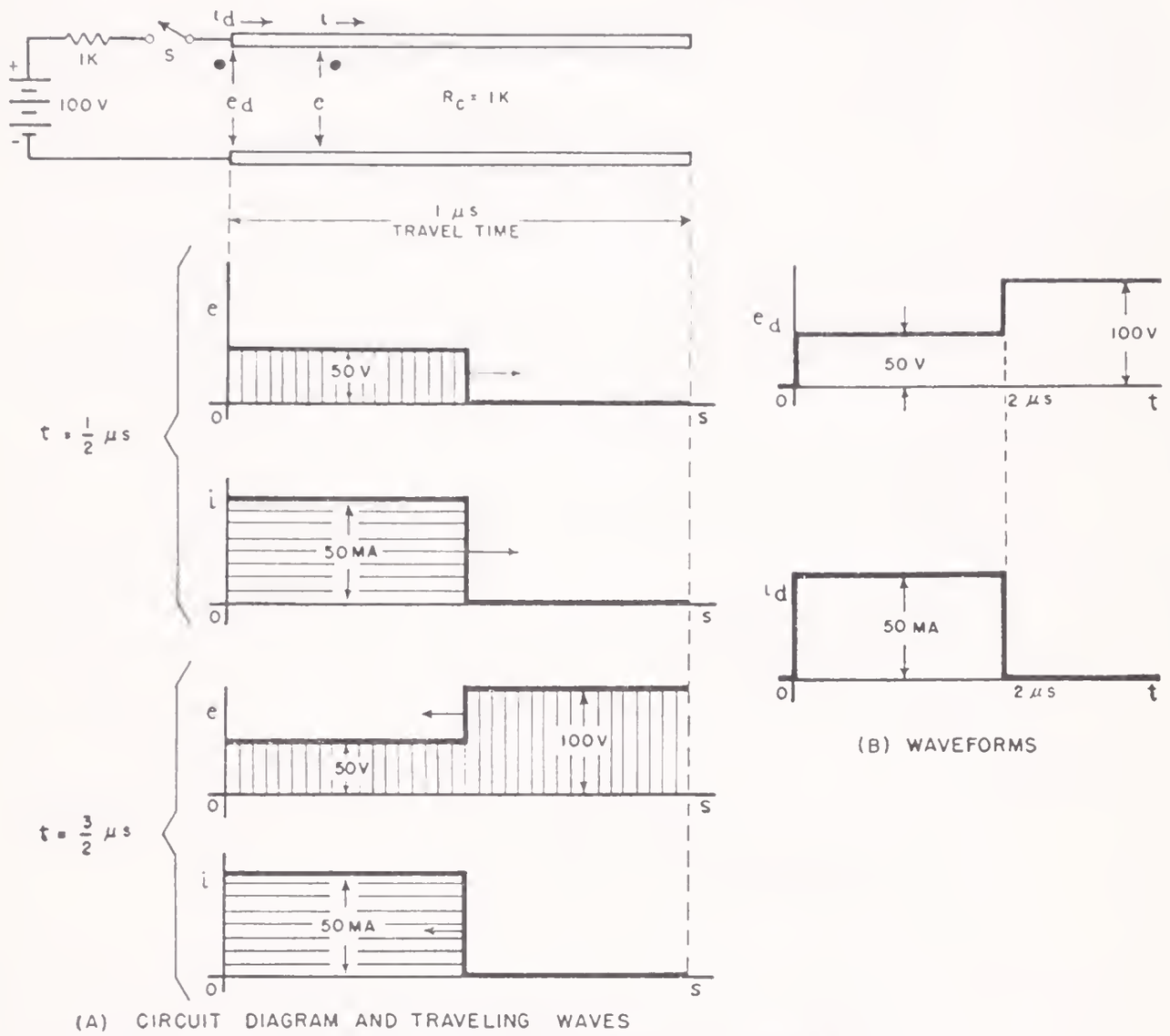


FIG. 58. Traveling waves on open-end line.

In the circuit of Fig. 59, a short circuit instead of an open circuit is employed at the far end of the transmission line. Consequently, the *voltage at the short circuit must always be zero*, and at the instant the 50-volt, 50-ma incident wave reaches the short circuit, a reflected voltage wave of 50-volt amplitude and reversed polarity starts toward the battery. A current wave of 50-ma amplitude is associated with the reflected voltage wave, and the current in this wave flows in the same direction as that in the incident wave. After the reflection, a current of 50-ma flows from the battery into the moving wavefront, and a current

of 100-ma flows from the wavefront to the short circuit. Thus the line capacitance at the wavefront is discharged by a net current of 50 ma, and the voltage between conductors drops from 50 volts to zero. When the reflected wave reaches the battery, the line voltage is zero at all points and the line current is 100 ma. No further change occurs because 100 ma is exactly the current drawn by a 1-K resistor connected across a 100-volt battery.

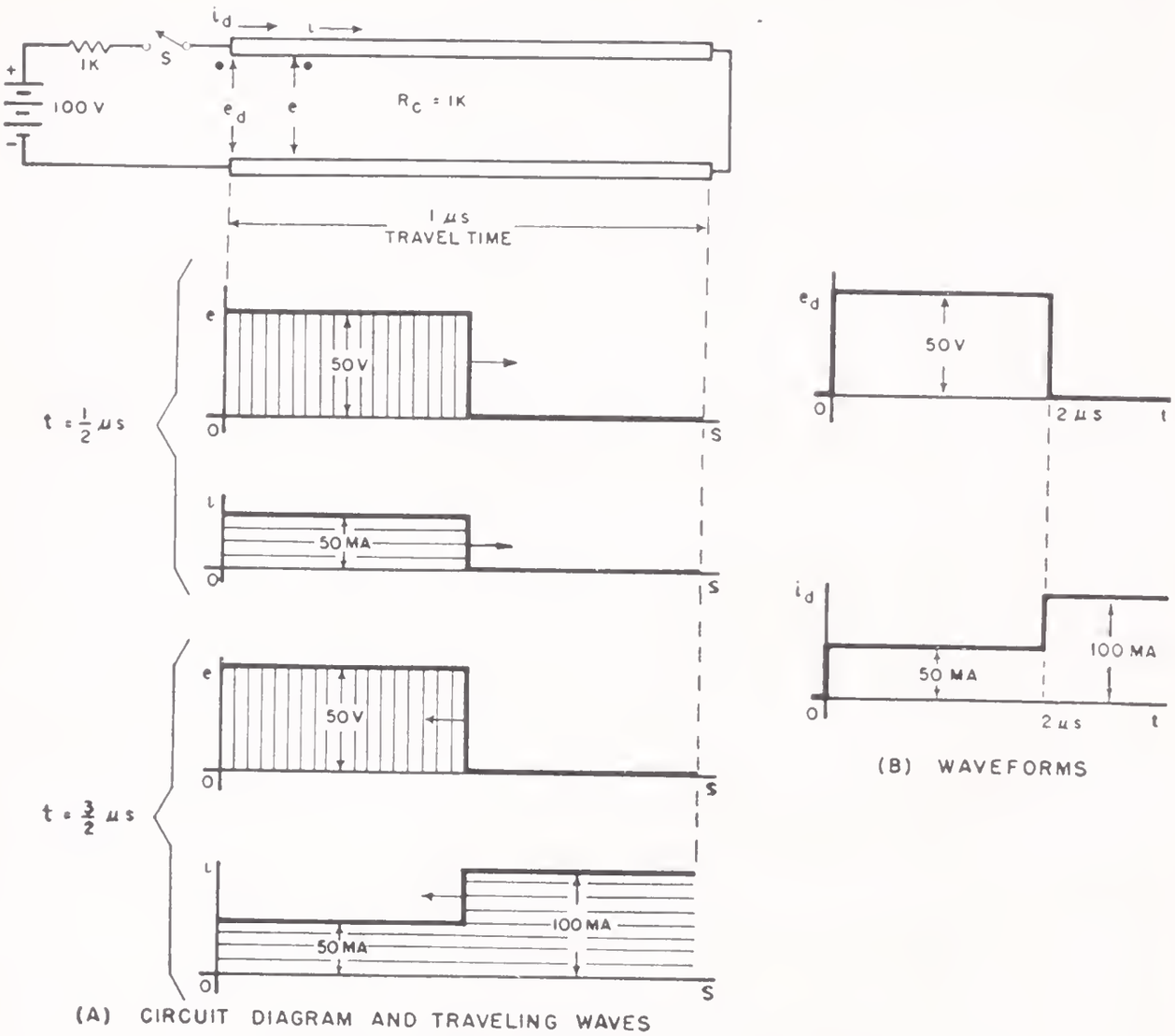


FIG. 59. Traveling waves on shorted-end line.

If a line is terminated in any manner whatsoever, the size and polarity of the reflected wave can be determined by use of the equivalent circuit developed in Art. 23. For example, in Fig. 60*A* a resistance of any value R_L terminates the line. A wave of voltage e^+ and current $i^+ = e^+/R_c$ travels toward the load end of the line after switch S is closed. When this wave reaches the load end of the line, conditions are the same as if R_L were suddenly connected to a line with initial voltage and current e^+ and i^+. The equivalent circuit of Fig. 60*B* is therefore applicable. The parallel combination of R_c and the current source in this circuit can be converted by Thévenin's theorem to R_c in series with a voltage source

$R_c i^+ = e^+$, and the values of the load current and voltage can then be calculated. The results are $2e^+/(R_L + R_c)$ for i_L and $2e^+R_L/(R_L + R_c)$ for e_L.

In Fig. 60*A*, the voltage of the wave incident at the load is e^+, and the sum of the voltages of the incident and reflected waves, $e^+ + e^-$, is the

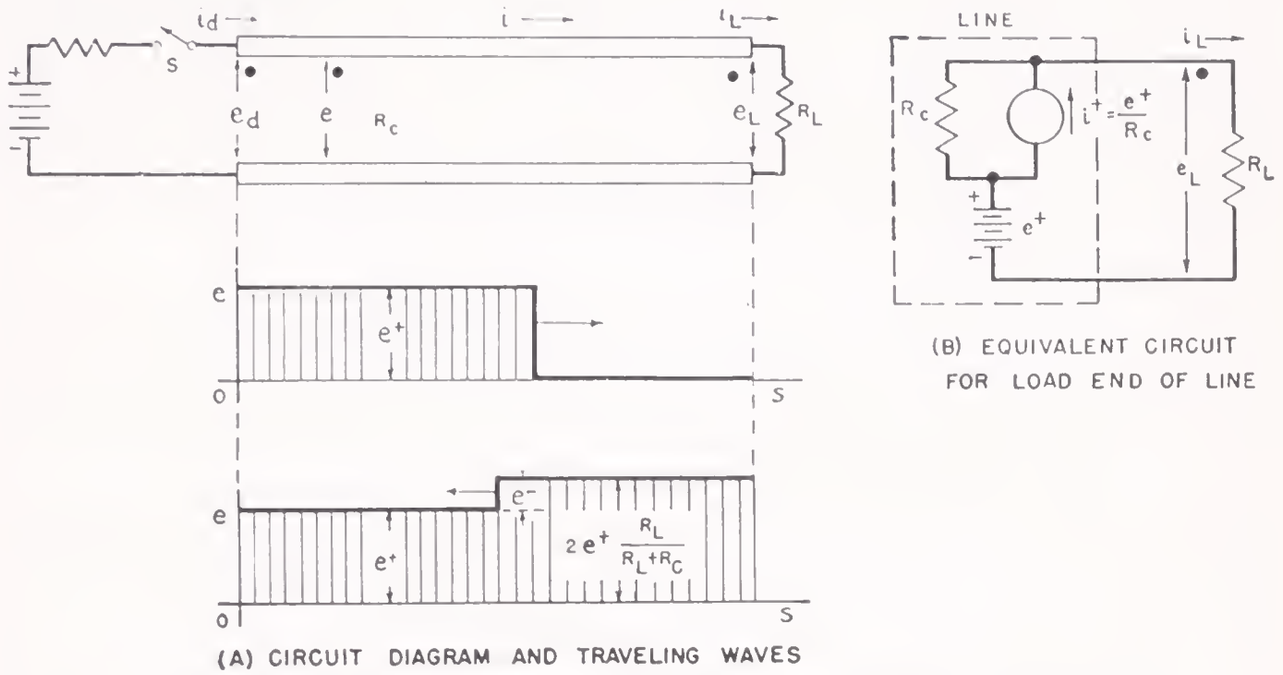


FIG. 60. Reflection caused by terminating resistance of any value.

total voltage $2e^+R_L/(R_L + R_c)$. Therefore, the reflected-wave voltage is

$$e^- = 2e^+ \frac{R_L}{R_L + R_c} - e^+ = e^+ \frac{R_L - R_c}{R_L + R_c}$$

The ratio of the reflected voltage wave to the incident voltage wave is called the *reflection coefficient* Γ. For the circuit of Fig. 60*A*

$$\Gamma = \frac{e^-}{e^+} = \frac{R_L - R_c}{R_L + R_c} \tag{41}$$

The reflection coefficient provides a quick method for determining the amplitude and polarity of the reflected wave, provided the line termination is a linear resistance. For example, if $R_L = R_c$ (as it does at the *battery* end of the lines in Figs. 58 and 59), $\Gamma = 0$ and no reflection occurs. If the line is terminated in a short circuit (as in Fig. 59), $R_L = 0$, $\Gamma = -1$, and a wave is reflected with full amplitude and reversed polarity.

The reflection coefficient applies to *changes* of voltage caused by a traveling wave. Therefore, initial voltages and currents of the line and battery voltages in the terminations should be ignored when using it. The reflection coefficient is correct only for linear-resistance terminations. If inductance, capacitance, or nonlinear elements (such as grid circuits in which current begins to flow during the rise of a wavefront) terminate

a transmission line, the basic equivalent-circuit method used in the derivation of the reflection coefficient should be used instead of the reflection coefficient.

25. Applications of Artificial Transmission Lines. The simplest use of an artificial line is to delay a waveform by the time required for it to traverse the line. Delays of the order of a microsecond are readily obtainable and are often useful in producing a desired sequence of operations in a radar system. Long delay times cannot conveniently be provided by electrical artificial lines, because of the tremendous numbers of sections required to provide sufficient delay and sufficiently high cutoff frequency.

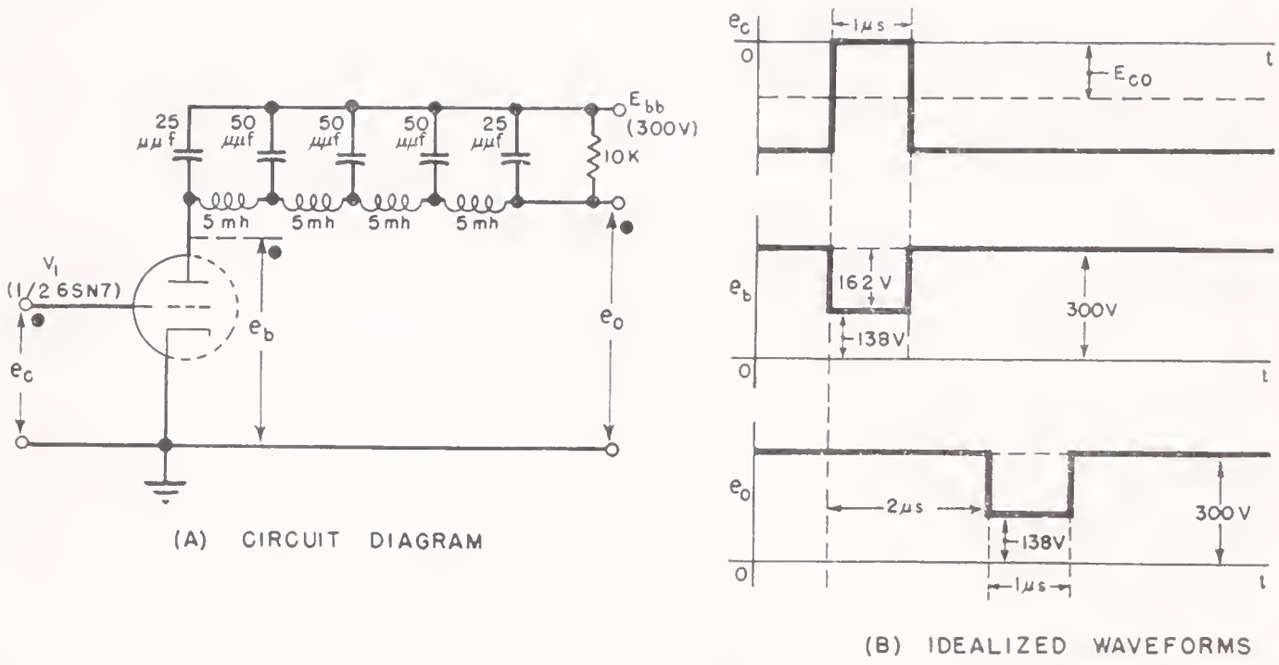


FIG. 61. Delay line.

A typical circuit employing a *delay line* is drawn in Fig. 61*A*. Capacitance values specified in this figure include the shunt capacitances of the circuit. (Tube and circuit capacitances often make up important parts of the capacitances at the ends of a delay line.) The characteristic resistance of the line is $R_c = \sqrt{(5 \times 10^3)/50} = 10$ K, and the total delay time is $4 \times \sqrt{(5 \times 10^{-3}) \times 50} = 2$ μs. Since the line is terminated in its characteristic resistance, pulses applied to the input end by V_1 travel to the output end and are not reflected, and operation of the tube is as though a 10-K resistor were connected into its plate circuit.

If tube V_1 is cutoff except during a 1-μs pulse applied to its grid, the plate voltage of e_b is a negative 1-μs pulse, as drawn in the idealized waveforms of Fig. 61*B*. The amplitude of this pulse is obtained from the plate characteristics of the tube and a 10-K, 300-volt load line. The pulse in the e_b wave appears also in the e_o wave, delayed 2 μs by the line. Sometimes a switch is provided for connecting the output terminals to

any one of the junctions along the line; in this manner a variable delay time is obtained. The load resistor is then left connected to the final output of the line, and the impedance of the switched circuit must be high relative to R_c to avoid reflections at the junction from which the output voltage is obtained.

A circuit that illustrates the use of an artificial line to discriminate among pulses of different durations is drawn in Fig. 62*A*. Tube V_1 is

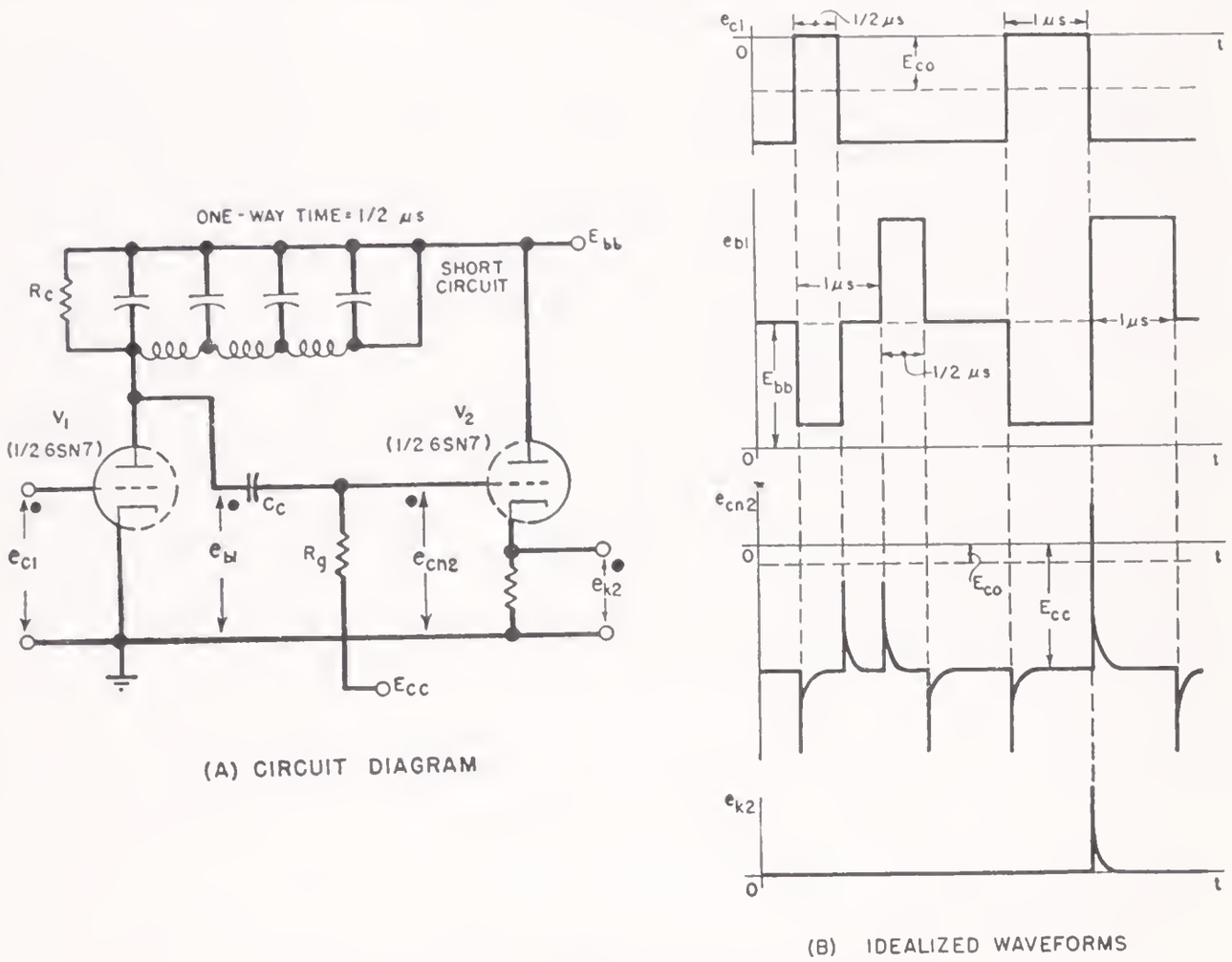


FIG. 62. Pulse-length-discrimination circuit.

normally cutoff but is made to conduct by a series of rectangular grid-voltage pulses of different durations. An output pulse is obtained from V_2 only when the input pulse has a duration of 1 μs.

In Fig. 62*B*, the effect of the circuit upon pulses of ½- and 1-μs duration is indicated. The ½-μs pulse causes a negative pulse of equal duration at the plate of V_1, followed after 1 μs by a *positive* pulse resulting from the reflection at the short-circuited end of the line. The positive pulse is not reflected at the tube end of the line because of the resistor R_c. (The tube is cutoff after the initial pulse, and R_g is so large relative to R_c that its conductance is negligible.) The C_c, R_g circuit has a very short time constant and peaks the e_{b1} waveform, as shown in the figure. Because of the large bias voltage for V_2, the peaks in the e_{cn2} wave caused

by the ½-μs input pulse are not of sufficiently large amplitude to cause V_2 to conduct.

The 1-μs input pulse in Fig. 62*B* also causes a negative and a positive pulse in the e_{b1} waveform. Since the pulse duration is equal to the two-way delay time of the line, the trailing edge of the negative pulse coincides in time with the leading edge of the positive pulse, and a double-amplitude peak appears in the e_{cn2} waveform. The bias for V_2 is set so that this double-amplitude peak, in contrast to the peaks caused by the ½-μs pulse, does cause a pulse in the output voltage e_{k2}.

Operation of the circuit with input pulses of more than 1-μs duration is somewhat more complicated because the reflected pulse reaches the input end of the artificial line while the tube is still conducting. The terminating resistance is then less than the characteristic resistance because of the shunt connection of the plate resistance of V_1, and additional reflections occur. If, however, the characteristic resistance of the line is small relative to the plate resistance of the tube, the amplitudes of the additional reflected pulses can be made too small to influence the output wave of V_2. Because the trailing edge of the initial pulse and the leading edge of the first reflected pulse coincide only for a pulse duration of 1 μs, only pulses of this one duration can produce an output pulse.

The most important use in radar of networks that simulate transmission lines is the formation of rectangular pulses. *Pulse-forming networks* are discussed in detail in Sec. A of the next chapter.

CHAPTER III

MODULATORS

A radar modulator, by controlling the operation of the transmitter, causes radio-frequency energy to be produced in short, powerful pulses. The modulator controls the shape, duration, and repetition frequency of the pulses. Usually it exercises control by applying pulses of plate-to-cathode voltage to the transmitter tube. (Magnetrons must always be plate modulated. Triode transmitter tubes are often plate modulated, but they may be controlled by pulses applied to the grid, or they may be self-pulsed.)

Nearly rectangular modulator-voltage pulses are preferred in most radar systems. A flat top is desired because both magnetrons and triode oscillators shift frequency if the supply voltage changes, and both operate at reduced efficiency if the supply voltage falls below its maximum. A steep leading edge is required in fire-control systems to make possible accurate measurement of range. A steep trailing edge is important when short minimum range is desired—the output obtained while the voltage is dropping from its maximum value to zero contributes little to the useful energy in the pulse, but it does obscure weak echoes from very close objects.

Short pulses (duration of 0.1 to 2.0 μs) required for high resolution are usually produced by means of pulse-forming lines or by blocking oscillators. Long pulses (durations of 2.0 to 20 μs) required to provide the pulse energy needed for long-range detection of small objects may be formed in a similar fashion, or they may be produced by multivibrators or by clipping and peaking circuits. In *line-pulsing modulators*, a high-power pulse is formed by a network and delivered directly to the transmitter. In *driver-power-amplifier modulators*, the pulse is formed in a low-power driver stage and is amplified before being applied to the transmitter.

Because the use of artificial-line networks is important in both line-pulsing and driver-power-amplifier modulators, Sec. A of this chapter is devoted to the principles of this method of pulse formation. The application of these principles and other details of actual modulator circuits are discussed in Sec. B for line-pulsing modulators and in Sec. C for driver-power-amplifier modulators.

A. PULSE FORMATION BY NETWORKS AND SWITCHES

Two types of pulse-forming networks are employed in modulators. One is the equivalent of an open-end transmission line. This type of network is charged relatively slowly in the time interval between pulses and then discharged suddenly to form the pulse. The other type of network is equivalent to a transmission line with a short circuit at one end. A current is built up in the line between pulses, and the branch of the external circuit that carries this current is suddenly opened when a pulse is desired. The voltage that appears across the line is the pulse.

1. Charging an Open-end Line. A circuit for charging an open-end transmission line (or its equivalent pulse-forming network) through a resistance is shown in Fig. 1*A*. The charging resistance is R_{ch}, and the switch S is closed to initiate the charging process. The form of the voltage at the input terminals of the line during charge depends upon the value of R_{ch} relative to R_c, the characteristic resistance of the network. Therefore, three charging processes are considered as follows:

Charging through a Resistance Equal to the Characteristic Resistance. If R_{ch} is equal to R_c, closing switch S causes the voltage e_d to rise to the battery voltage in two equal steps, indicated by the curve for $R_{ch} = R_c$ in Fig. 1*C*, as explained in Art. 24, Chap. II. The time T required for e_d to reach its final value is *twice* the time for a wave to travel one way along the line.

Charging through a Resistance Greater than the Characteristic Resistance. If R_{ch} is greater than R_c, more than two steps are required to complete the charging of the network. As an example, suppose $R_{ch} = 9R_c$. When S is first closed, the equivalent circuit (see Art. 23, Chap. II) is that of Fig. 1*B* with $E_{st} = 0$; and the initial value of e_d is

$$E_{bb} \frac{R_c}{R_c + 9R_c} = 0.1E_{bb}$$

as plotted in Fig. 1*C*. A voltage wavefront of height $0.1E_{bb}$ traverses the line toward the open circuit, where it is reflected without change in amplitude or polarity and returns to the input terminals, as indicated in Fig. 2*A*. At the time $t = T$, the entire line is charged to the voltage $0.2E_{bb}$, and the current at every point in the line is zero. Thus, if the switch S were opened at this instant, the voltage at every point on the line would remain at $0.2E_{bb}$, the voltage indicated by point A in Fig. 1*C*.

The switch is not opened. Therefore, current from the battery flows into the line to start a new wavefront. The voltage of the second wavefront is found from Fig. 1*B* with $E_{st} = 0.2E_{bb}$, the voltage to which the

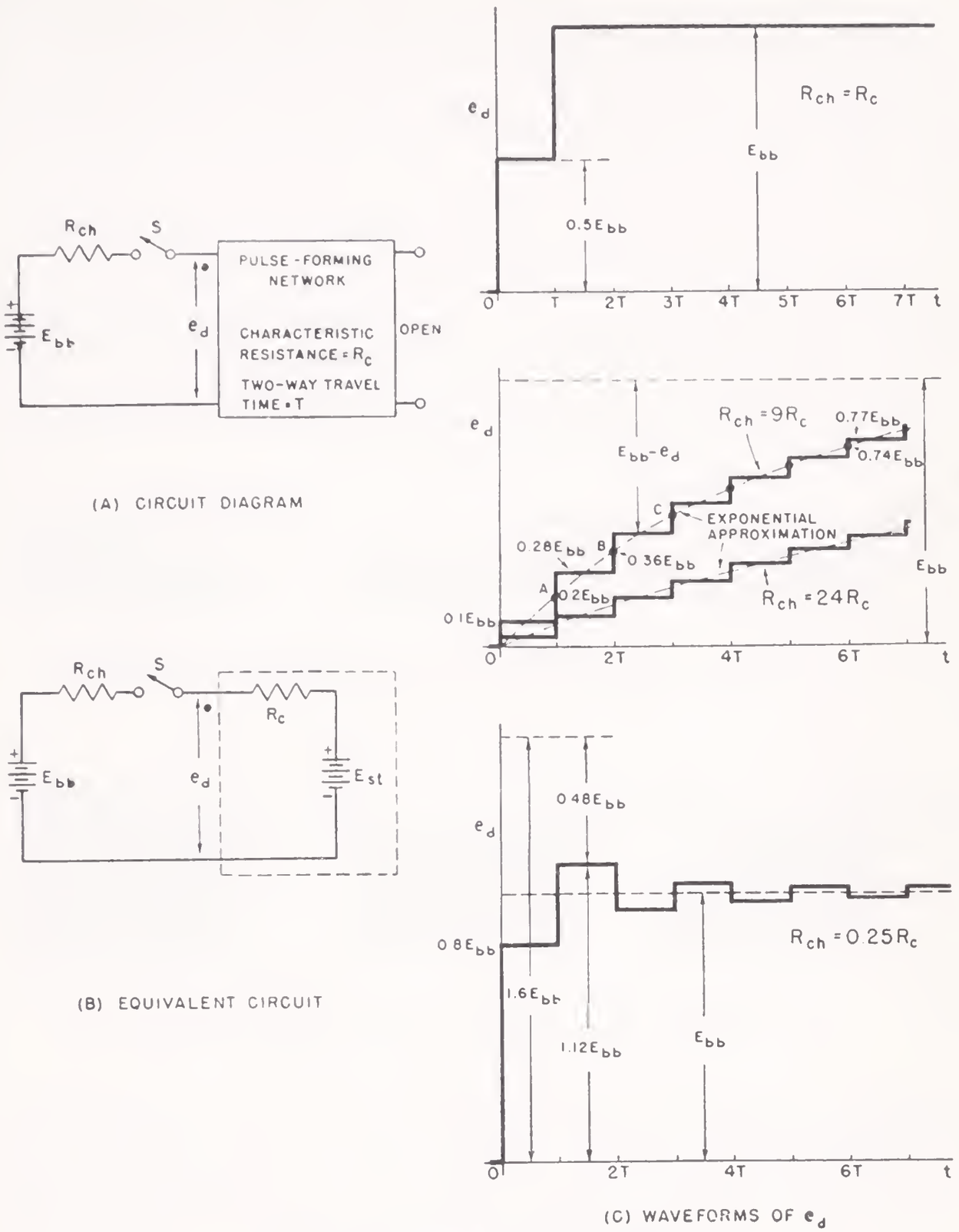


FIG. 1. Charging of open-end line.

line is charged. The second wavefront, of height

$$\frac{R_c}{R_c + 9R_c}\,(E_{bb} - 0.2E_{bb}) = 0.08E_{bb}$$

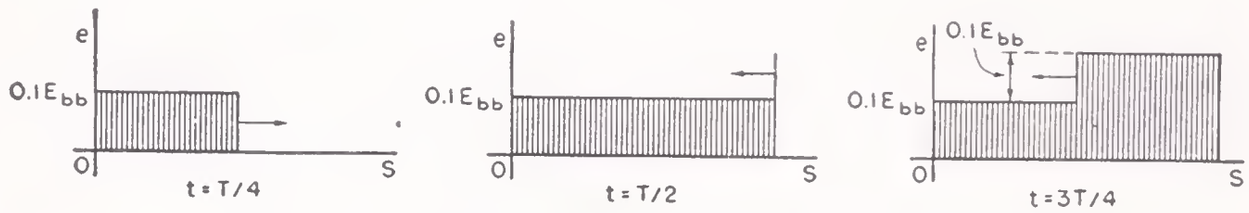
traverses the line, is reflected, and returns to the input, as indicated in Fig. 2*B*. By continuing this reasoning, the voltages of all the steps of the curve for $R_{ch} = 9R_c$ in Fig. 1*C* may be found.

An alternate method of finding the height of the wavefronts reflected from the input end of the line makes use of the reflection coefficient. For example, at time $t = T$ the wavefront arriving at the battery end of the line is $0.1E_{bb}$. The reflection coefficient (see Art. 24, Chap. II) for $R_{ch} = 9R_c$ is

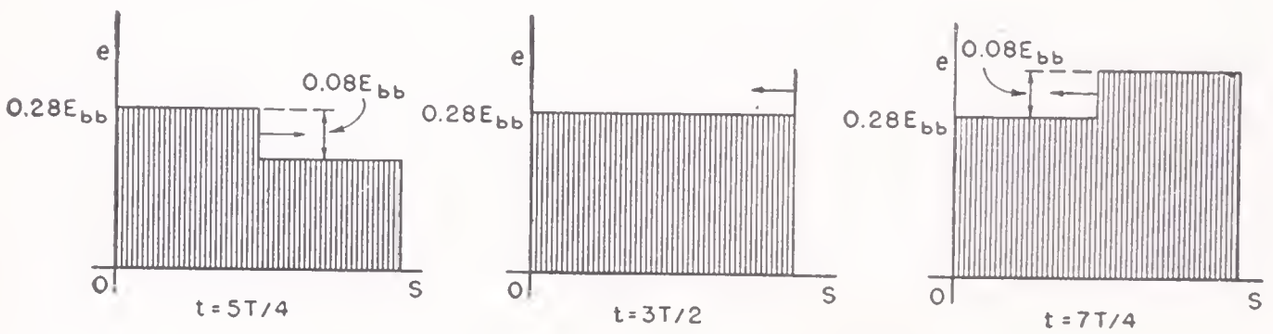
$$\Gamma = \frac{9 - 1}{9 + 1} = +0.8$$

Thus the reflected wave is $0.8 \times 0.1E_{bb} = 0.08E_{bb}$, which value agrees with the result obtained from the equivalent circuit.

Values of R_{ch} greater than R_c cause the network to require more than two steps for charging. The higher the value of R_{ch}, the smaller is each



(A) FIRST WAVEFRONT OF HEIGHT $0.1E_{bb}$ TRAVELS OUT AND BACK



(B) SECOND WAVEFRONT OF HEIGHT $0.08E_{bb}$ TRAVELS OUT AND BACK

FIG. 2. Traveling wavefronts on artificial line when $R_{ch} = 9R_c$.

step and the longer the time required to charge to a given percentage of E_{bb}. (Compare curves for $R_{ch} = 24R_c$ and $R_{ch} = 9R_c$ in Fig. 1*C*.) The dashed curve in Fig. 1*C*, which is drawn through the origin and points A, B, C, . . . in the waveform of e_d, is an exponential. Point A represents the voltage to which the entire line is charged after the first wavefront ($0.1E_{bb}$) has completed its round trip; point B represents the line voltage after the second wavefront ($0.08E_{bb}$) has traveled to the open end and returned; and so on. If R_{ch} is made very large compared with the characteristic resistance, the curve of e_d consists of very small variations above and below the exponential curve, and a very good approximation is had by assuming e_d to be the exponential. Furthermore, if R_{ch} is large compared with R_c, the time constant of the exponential is $R_{ch}C_{st}$, where C_{st} is the storage capacitance of the network, that is, the capacitance

measured with all inductances short-circuited.[1] In a slow charging process, the network is equivalent to a capacitance equal to C_{st} because the voltages across inductances are small compared with the voltage across the charging resistance, and these small voltages have little effect upon the current.

Charging through a Resistance Smaller than the Characteristic Resistance. If R_{ch} is smaller than R_c, the line voltage rises above E_{bb} at the end of the first step and thereafter oscillates in steps above and below E_{bb}. For example, suppose that, in Fig. 1*A*, $R_{ch} = 0.25R_c$. The equivalent circuit shows the initial value of e_d to be

$$\frac{R_c}{R_{ch} + 0.25R_{ch}} E_{bb} = 0.8E_{bb}$$

as plotted in Fig. 1*C*. A wavefront of this voltage travels to the open end, is reflected, and returns. Hence, at the time $t = T$, the line is charged throughout to $1.6E_{bb}$, which is higher than the battery voltage.

[1] The dashed curve of Fig. 1*C* can be shown to be an exponential as follows: At point *A*, corresponding to $t = T$, the voltage $E_{bb} - e_d$ is

$$E_{bb} - \frac{2R_c}{R_c + R_{ch}} E_{bb} = \Gamma E_{bb}$$

This voltage ΓE_{bb} plays the same role in the second step as E_{bb} did in the first. Therefore at point *B*, $t = 2T$, and $E_{bb} - e_d = \Gamma^2 E_{bb}$. Correspondingly, at the point on the dotted curve corresponding to $t = nT$, the voltage $E_{bb} - e_d$ is $\Gamma^n E_{bb}$. Thus, writing $n = t/T$ yields the equation of the dotted curve:

$$e_d = E_{bb} - \Gamma^{t/T} E_{bb} = E_{bb}(1 - \Gamma^{t/T})$$

This is the equation of an exponential. It may be written to the base ϵ by making the replacement $\Gamma = \epsilon^{\ln \Gamma}$. Thus,

$$e_d = E_{bb}(1 - \epsilon^{t \ln \Gamma / T}) \simeq E_{bb}[1 - \epsilon^{-t \ln (1/\Gamma)/T}]$$

and the time constant of the exponential is $T/(\ln 1/\Gamma)$.

If the pulse-forming network is an artificial line, proof that this time constant is $R_{ch}C_{st}$ for R_{ch} much greater than R_c is as follows: The pulse duration is $T = 2\sqrt{L_t C_{st}}$ (where L_t is the total line inductance), and the characteristic resistance is $R_c = \sqrt{L_t/C_{st}}$. If R_{ch} is much greater than R_c,

$$\frac{1}{\Gamma} = 1 + \frac{2R_c}{R_{ch} - R_c} \approx 1 + 2\frac{R_c}{R_{ch}}$$

and

$$\ln \frac{1}{\Gamma} \approx \ln \left(1 + \frac{2R_c}{R_{ch}}\right) \approx \frac{2R_c}{R_{ch}}$$

The last approximation is equivalent to replacing $\epsilon^{2R_c/R_{ch}}$ by $1 + 2R_c/R_{ch}$, which is allowable for small values of $2R_c/R_{ch}$, since the higher power terms of the series expansion of the exponential are negligible. Thus the time constant $T/(\ln 1/\Gamma)$ is

$$\frac{T}{\ln \frac{1}{\Gamma}} = \frac{T}{2R_c/R_{ch}} = \frac{2\sqrt{L_t C_{st}}}{2\sqrt{L_t/C_{st}}} R_{ch} = R_{ch}C_{st}$$

The next wavefront round trip involves the *discharge* of the line—the line current reverses, and a negative wavefront (that is, a change from a higher to a lower voltage) travels along the line. The magnitude of this and succeeding wavefronts may be calculated from the equivalent circuit or the reflection coefficient. At each reflection a wave somewhat larger than that necessary to restore e_d to E_{bb} is started along the line; therefore

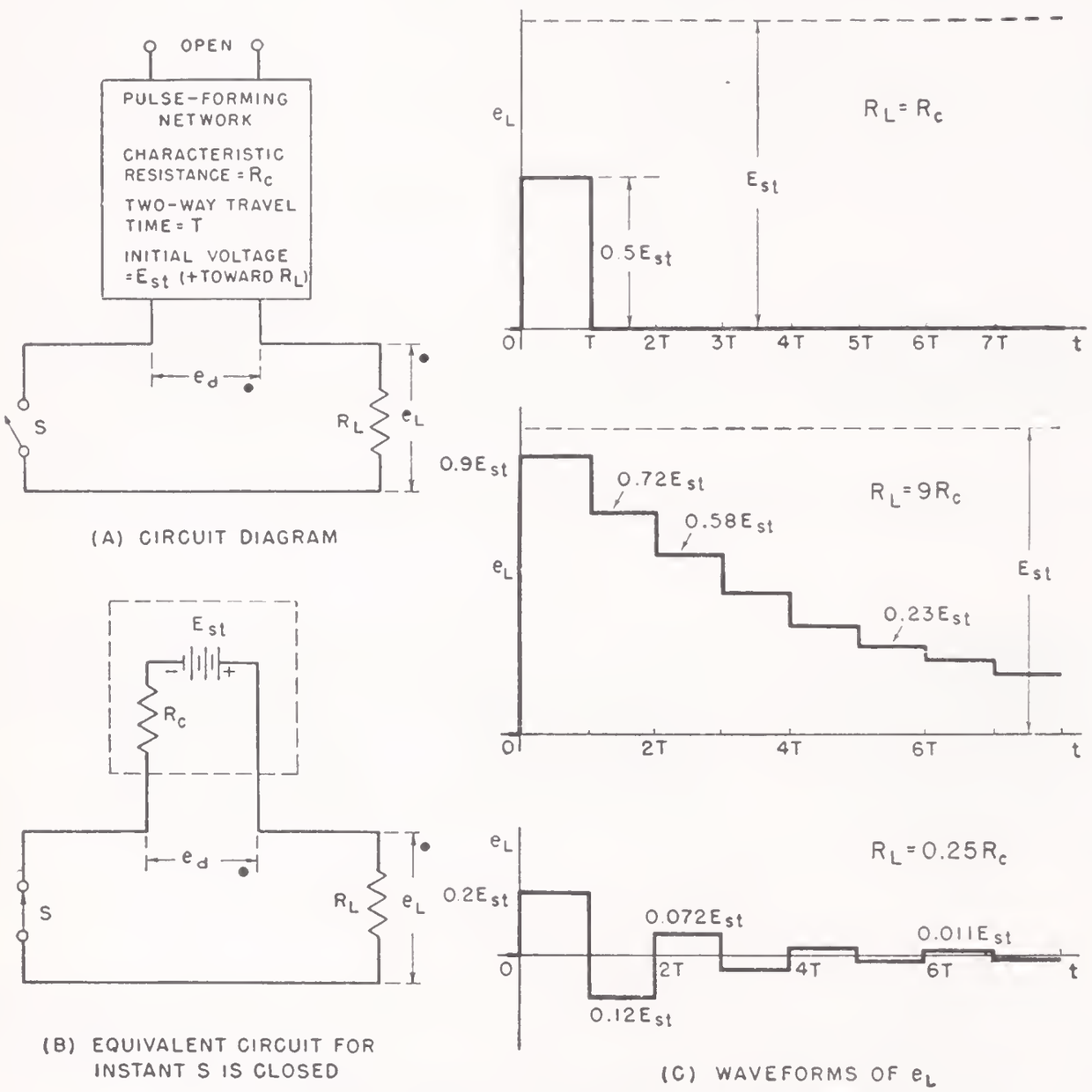


FIG. 3. Pulse formation by discharge of an open-end line.

the voltage e_d varies above and below E_{bb} in a damped square-wave oscillation.

2. Pulse Formation by Discharge of an Open-end Line. To produce a pulse by means of an open-end-line type of network, the network is discharged through a resistance, as in Fig. 3*A*. In this figure, switch S is closed to produce the discharge. The form of e_L, the voltage across resistor R_L, depends upon the relative sizes of R_L and R_c, as indicated by the curves in Fig. 3*C*. Observe the similarity of the curves in Figs. 3*C* and 1*C*.

The curve of e_L obtained with $R_L = R_c$ is determined as follows: Let E_{st} be the initial voltage to which the network is charged. Then, from the equivalent circuit of Fig. 3*B*, the voltage e_L is found to rise to $E_{st}/2$ when S is first closed; the line voltage e_d drops from E_{st} to $E_{st}/2$ at the same time. A negative wavefront progresses to the open end (leaving the line charged to $E_{st}/2$ at each point) and returns to the input end (leaving the line completely discharged). At this instant not only the line voltage but also the line current is zero at every point. Thus the energy originally stored in the line capacitors is completely transferred to the resistance. The rectangular pulse of voltage produced across R_L has a magnitude $E_{st}/2$ and a duration T equal to the *two-way* travel time.

Curves of e_L obtained with other values of R_I may be found in a similar fashion. The waveforms of Fig. 3*C* show that a load resistance greater than R_c causes e_L to remain positive after the end of the main part of the pulse, while a value of R_L less than R_c causes an overshoot and several smaller pulses following the first. Usually R_L is made as nearly as possible equal to R_c in order to obtain a single rectangular pulse. Sometimes R_L is purposely made less, because overshoot is desired with certain types of switches. The smaller pulses following the main one are then eliminated because the switch opens during the first overshoot.

3. Pulse Formation by a Shorted-end Line. An artificial line having one pair of terminals connected by a short circuit, or the equivalent pulse-forming network, can be used to form a pulse of voltage at the other pair of terminals. The network is prepared for the pulse-forming process by circulating a direct current I through it, as shown in Fig. 4*A* (switch S is initially closed). The switch is suddenly opened, and the voltage that appears across the load resistor R_L is the pulse.

Figure 4*B* serves as an equivalent circuit for the network when S is closed and for the first instant after it is opened. Since the voltage e_L is zero during the time that S is closed and a steady current I is flowing through the line, the equivalent circuit of the line comprises a resistor R_c shunted by a current source I. When S is first opened, the current I flows through resistors R_c and R_L in parallel and produces a voltage e_L.

If, for example, $R_L = R_c$, e_L rises suddenly to $\frac{1}{2} IR_c$ when switch S is opened. Thus a voltage wavefront of this amount progresses to the short-circuited end of the line, is reflected *with change of polarity*, and returns to the input terminals. The waveform of e_L is shown by the curve for $R_L = R_c$ in Fig. 4*C*. The initial wavefront reduces the current at each point of the line from I to $I/2$, and the returning wavefront further reduces the line current to zero. At the time $t = T$, therefore, line voltage and line current are both zero, and the energy originally stored in the line inductances is completely transferred to the load resistor.

The same methods may be applied to determine the other waveforms in Fig. 4*C*. The value of the current I in Fig. 4*B* is changed after each round trip made by a wavefront in a manner similar to the change of E_{st} in Fig. 3*B*. Alternatively, the reflection coefficient may be used to compute the amplitude of each wavefront after the first. Note that the

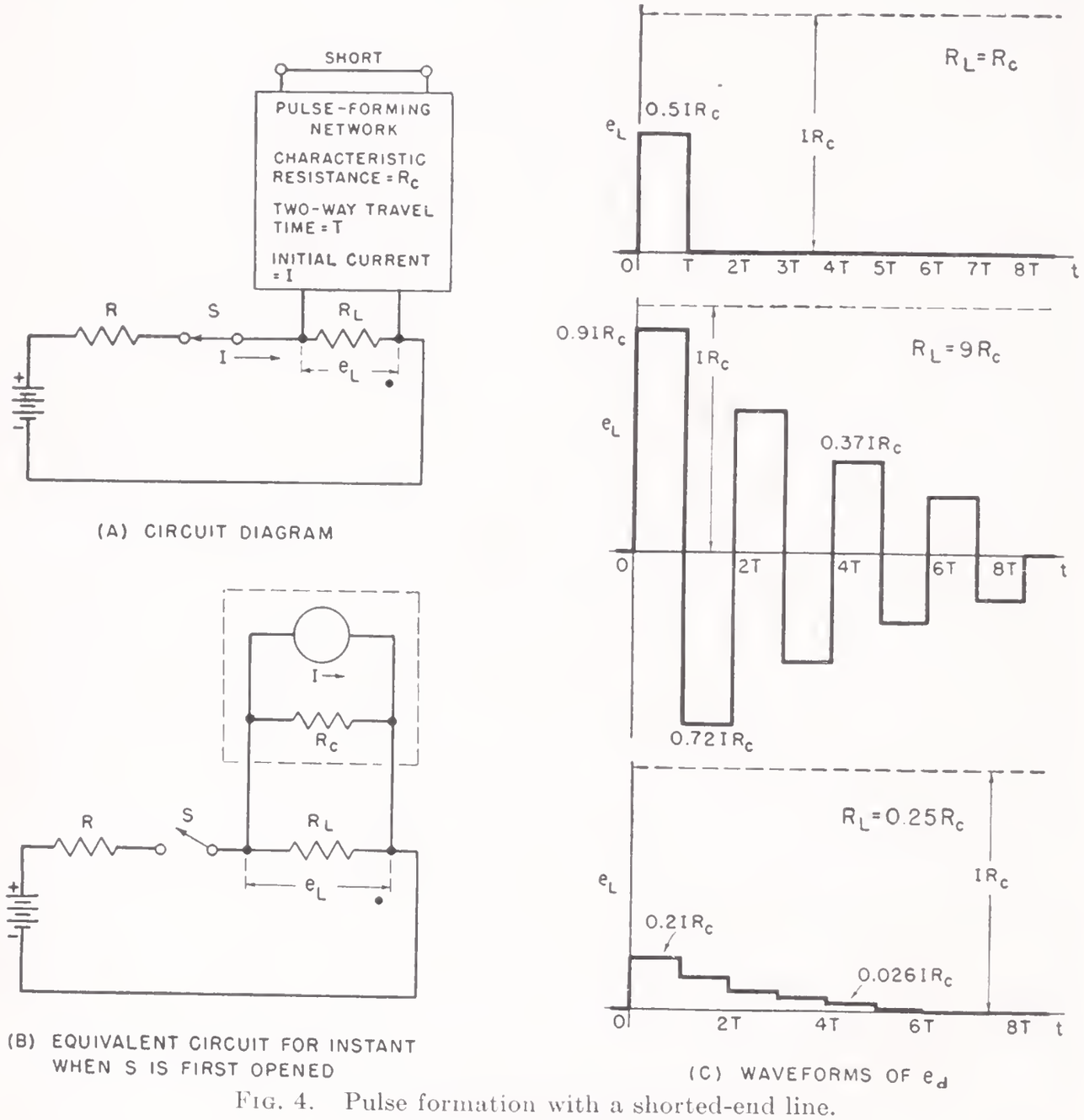


FIG. 4. Pulse formation with a shorted-end line.

voltage waveform obtained with R_L less than R_c is similar to that for the open-end line with R_L greater than R_c, and vice versa.

4. Pulse-forming Networks. Ordinary artificial lines, such as those in Fig. 52 of Chap. II, p. 139, are often employed for pulse formation in driver stages of modulators. Small imperfections of pulse shape are caused by the failure of the network to simulate a real transmission line perfectly, but these imperfections can frequently be removed by clipping in the following amplifier stages. Furthermore, the network operates at

relatively low voltage; thus a large number of sections may be used, if necessary, without making the physical size of the network unreasonable.

In line-pulsing modulators, the pulse-forming network must operate at high voltage and at the same time must provide a pulse of almost perfectly rectangular shape to be used by the transmitter. Consequently, a network that produces an acceptable pulse with the minimum number of elements is required.

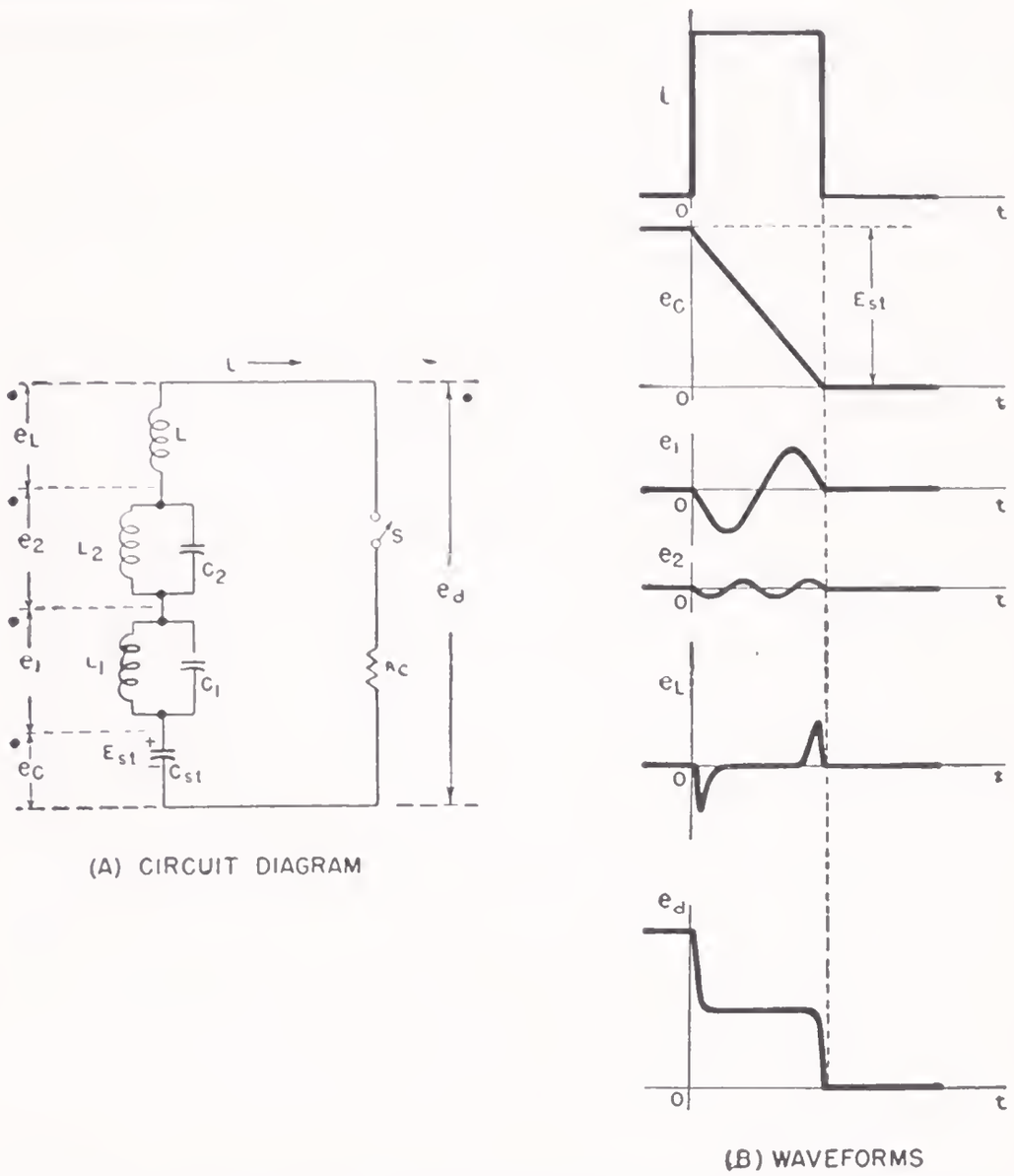


FIG. 5. Development of pulse voltage by pulse-forming network.

One network often used instead of an open-end artificial line is illustrated in Fig. 5*A*. If C_{st} is charged to a high voltage E_{st} and then switch S is closed so that the network is discharged through its characteristic resistance R_c, nearly rectangular voltage and current pulses are obtained across R_c. Oscillatory voltages across the L-C circuits add to the voltage of C_{st} during discharge in such a manner that the network voltage falls to zero in two equal steps.

The curves of Fig. 5*B* show how this result is accomplished. The current waveform represents the rectangular pulse of current assumed to be

obtained when the switch S is closed. The constant current of the pulse discharges capacitor C_{st} linearly, giving rise to the capacitor-voltage waveform e_C. The size of the capacitor is chosen so that zero voltage is reached at the end of the pulse. The front edge of the pulse of current starts an oscillation in the L_1, C_1 circuit (see Art. 18, Chap. II), and the end of the pulse of current ends the oscillation by starting another of equal amplitude but opposite polarity. The values of L_1 and C_1 are chosen so that 1 cycle of the oscillation occurs during the pulse, as shown by the waveform of e_1 in Fig. 5*B*. An oscillation is started and stopped in L_2 and C_2 in a similar way, but the frequency of this oscillation is higher, and 2 cycles of it occur during the pulse. If the current pulse were perfectly rectangular, infinite-amplitude peaks of voltage would appear across the inductance L. Actually, of course, with this inductance in the circuit the edges of the current pulse must slope a little, and finite voltage peaks are produced, as shown by the waveform of e_L. The inductance L is small, and the slope of the edges of the current pulse is not sufficient to affect the other voltage waveforms appreciably.

Adding together the voltage waveforms e_C, e_1, e_2, and e_L yields the network terminal voltage e_d. Because the total voltage has a two-step form, a nearly rectangular voltage pulse is applied to the resistance R_c when switch S is closed, and the original assumption of a rectangular current pulse is justified.

By varying the values of the inductances and capacitances in the resonant circuits, the amplitudes and frequencies of the oscillations added to the capacitor voltage can be controlled. Increasing the amplitude of the higher frequency terms increases the steepness of the sides of the pulse but causes oscillations to appear following the edges of the pulse. Thus a means of obtaining the best compromise between flat top and steep edges is available; such a means of adjustment is not available with the ordinary artificial line.

Some other types of pulse-forming networks are shown in Fig. 6. The first network in Fig. 6*A* is used when an appreciable deviation of the voltage pulse from a rectangular shape is allowable. The second network in Fig. 6*A* is equivalent in performance to the one of Fig. 5*A*. A network that simulates a short-circuited transmission line is shown in Fig. 6*B*. Energy is stored in the magnetic field of the inductance L_{st} because of an initial current, and when current in a part of the external circuit is interrupted, a voltage pulse is produced across L_{st}. The elements shunted across L_{st} control the shape of this pulse and make it nearly rectangular.

The network in Fig. 6*C* is an open-end artificial line with magnetic-field coupling between the line inductances. Without mutual inductance, an artificial line can provide pulses with very steep edges, but undesirable oscillations appear on the top of the pulses. Correct adjustment of the

mutual inductance allows the oscillations to be removed with only a small decrease in the steepness of the edges.[1] The artificial line with mutual inductance is usually constructed by winding a long solenoid and placing tap points at equal intervals along the winding. Equal-valued capacitors are connected to the tap points. The combination of all capacitors in parallel is the storage capacitance. This type of network is the one most frequently used when an accurately rectangular pulse is desired.

Pulse-forming networks are often enclosed in metal boxes with each of the circuit elements immersed in oil to give compact high-voltage insula-

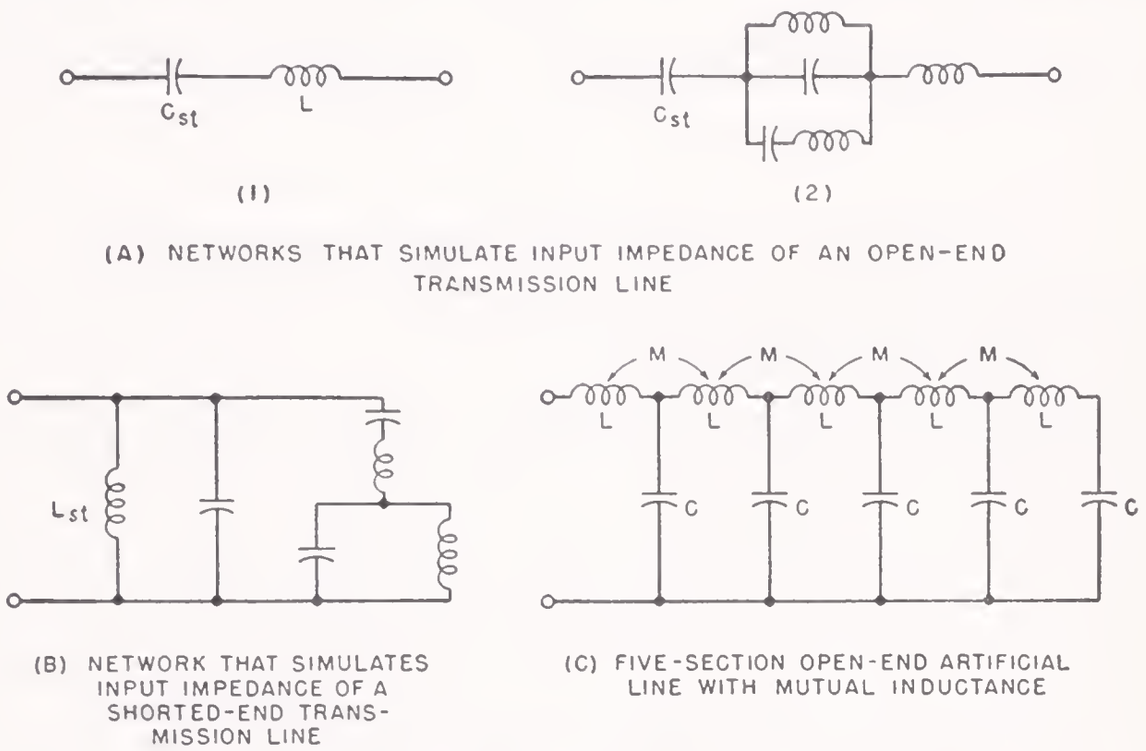


FIG. 6. Pulse-forming networks used in radar modulators.

tion. If one element fails, the entire network must be replaced. Networks are labeled with the value of their pulse length T, characteristic resistance R_c, and safe operating voltage.

If the network is of the open-end-line type, the value of the storage capacitance C_{st} can be calculated from the values of T and R_c. Suppose a lossless network is charged to a voltage E_{st} and then discharged into its characteristic resistance. A voltage pulse of duration T and amplitude $E_{st}/2$ is produced. Since the energy originally stored in the capacitance C_{st} is delivered to the load resistance R_c during discharge,

$$\frac{1}{2} C_{st} E_{st}^2 = \left(\frac{E_{st}}{2}\right)\left(\frac{E_{st}}{2R_c}\right) T$$

[1] The design of artificial lines with mutual inductance and of pulse-forming networks like those of Fig. 6*A* is discussed in E. A. GUILLEMIN, "A Historical Account of the Development of a Design Procedure for Pulse-Forming Networks," Report 43–10/16/44 (M.I.T. Radiation Laboratory, Cambridge, Mass., 1944).

or

$$C_{st} = \frac{T}{2R_c}$$

and

$$T = 2R_cC_{st} \tag{1}$$

Equation 1 shows that the pulse duration equals two time constants of a circuit composed of the characteristic resistance and storage capacitance of the network.

If the network is equivalent to a shorted-end line, the storage inductance L_{st} may be shown in a similar way to be related to T and R_c by the equation

$$T = 2\frac{L_{st}}{R_c} \tag{2}$$

That is, the pulse duration of a shorted-end-line type of network equals two time constants of the circuit comprising the characteristic resistance and the storage inductance.

5. Characteristics of Loads for Pulse-forming Networks. In Arts. 2 and 3, a pulse-forming network is assumed to be connected to a linear load resistance R_L during the generation of a pulse. In actual circuits, the network is connected to a nonlinear vacuum-tube circuit—the plate circuit of a magnetron or triode transmitting tube in line-pulsing modulators and the grid circuit of an amplifier tube in driver-power-amplifier modulators. Neither magnetrons nor triode oscillators draw a plate current proportional to the plate voltage. Amplifier grid circuits are nonlinear because the grid voltage changes from a negative value before the pulse to a positive value during the pulse.

Despite the nonlinearities of these loads, the principles developed in Arts. 2 and 3 may be applied to actual circuits, provided Thévenin's theorem is used (rather than the reflection coefficient) in calculating the voltage produced when a traveling wave reaches a nonlinear load at the end of an artificial line.

Consider, as an example, the pulse-forming network of Fig. 7*A* which is charged to 36 kv and then discharged through the magnetron. As explained in Art. 8, Chap. XI, the plate current of a magnetron rises slowly as plate voltage is increased until a critical value of voltage is reached and then rises very rapidly for further increase of voltage. Such an i_b, e_b characteristic is drawn in Fig. 7*C* and is markedly nonlinear.

When switch S is first closed, the Thévenin-theorem equivalent of the network is a 36-kv battery in series with a 1.2-K resistor, as in Fig. 7*B*. If the operating point of the magnetron lies somewhere on the straight line AB of Fig. 7*C*, the magnetron is equivalent to a 15-kv battery in

series with a 200-ohm resistor representing the incremental plate resistance r_p, and the circuit of Fig. 7*B* results. (Compare the magnetron equivalent circuit with the triode equivalent plate circuit of Chap. II. See p. 60.)

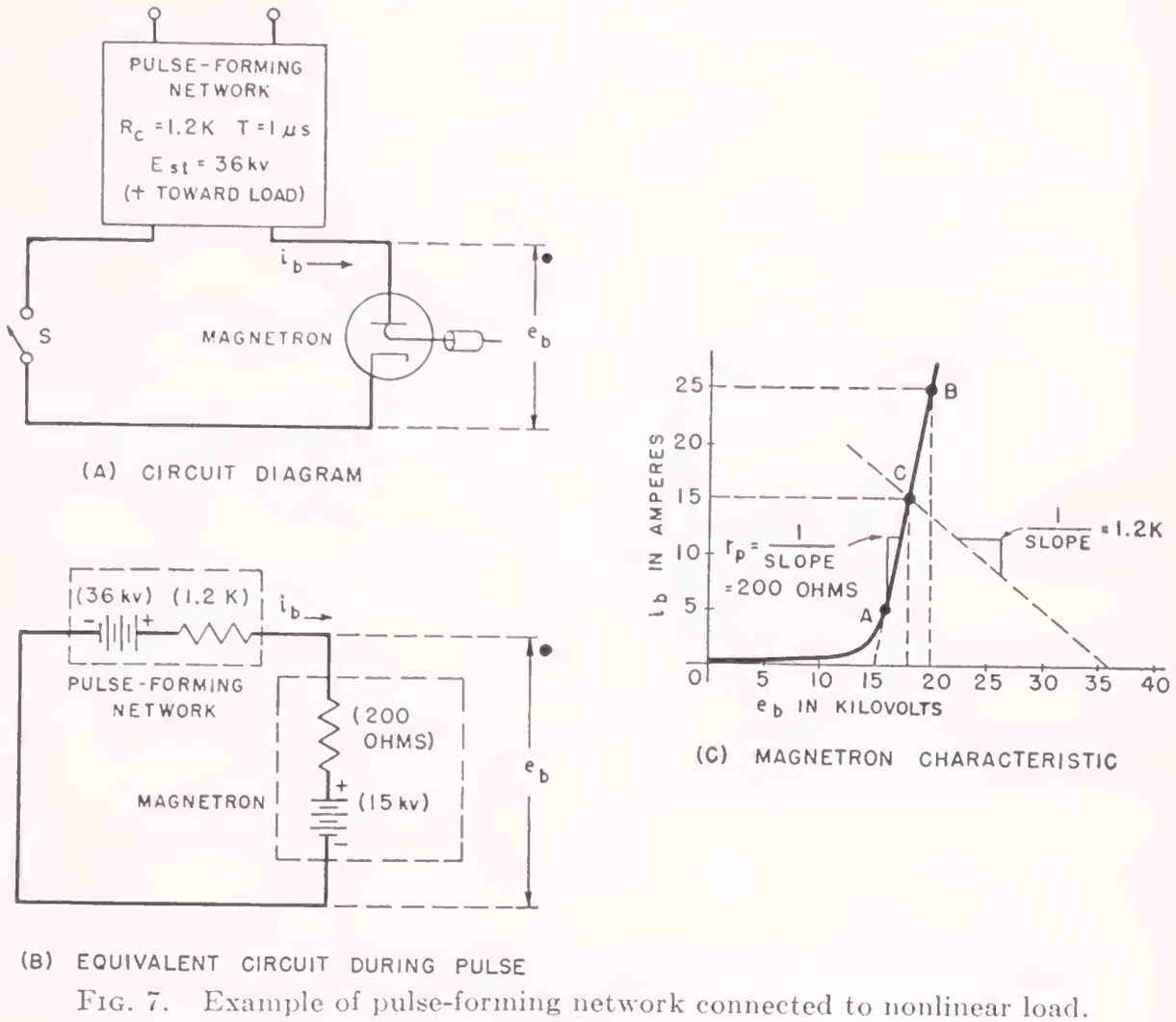


FIG. 7. Example of pulse-forming network connected to nonlinear load.

From Fig. 7*B*, the magnetron current during the pulse is found to be

$$i_b = \frac{36{,}000 - 15{,}000}{1{,}200 + 200} = 15 \text{ amp}$$

The magnetron voltage is therefore

$$e_b = 15{,}000 + 200 \times 15 = 18{,}000 \text{ volts}$$

and the operating point is *C* in Fig. 7*C*. This point can be located in another way without the use of an equivalent circuit for the magnetron: The pulse-forming network can be considered as a plate-voltage source of 36 kv in series with a 1.2-K load resistor. The intersection of the load line for this source (shown dashed in Fig. 7*C*) with the magnetron characteristic locates point *C*.

During the pulse, the ratio e_b/i_b is 18,000/15 = 1,200 ohms. Thus the network operates as if it were connected to its characteristic resistance

and delivers a single 1-μs pulse of voltage equal to one-half of E_{st}. This matched condition results, however, only for the particular value of E_{st} assumed. If E_{st} were reduced, plate current would decrease more than plate voltage and the network would appear to discharge into more than its characteristic resistance. For example, if E_{st} were 29 kv, i_b would be 10 amp, according to Fig. 7*C*, and e_b would be 17 kv. The ratio $e_b/i_b = 1.7$ K is more than R_c; the initial pulse then has an amplitude more than one-half of E_{st} and is followed by a series of smaller pulses. If E_{st} were greater than 36 kv, the network would effectively be connected to less than its characteristic resistance and an initial low-amplitude pulse would be followed (if S remained closed) by alternately negative and positive pulses. Note that the steep slope of the i_b, e_b curve of the magnetron causes a large change in the amplitude of the current pulse to be associated with even a small change in the amplitude of the voltage pulse.

6. Switching Devices for Discharging Pulse-forming Networks. The pulse-forming network in a modulator must be discharged through some form of switching device that is closed at the beginning of a pulse and opened soon after the end of the pulse to allow the storage elements of the network to be recharged. The characteristics of a device that affect its usefulness as a switch for pulse-forming networks are (1) the precision with which the firing time can be controlled, (2) the ability of the device to conduct current until the network is discharged and to become an open circuit after the network is discharged, (3) the voltage drop across the device when it is conducting, (4) the peak voltage at which the device can successfully operate, (5) the life of the device under operating conditions, and (6) the influence of such variables as pressure, temperature, humidity, and age on the performance of the device.

Commonly used switching devices are discussed in the following paragraphs, and a summary of their characteristics is given in Table 1.

Rotary Spark Gaps. A rotary spark gap (see Fig. 8) is an effective switch because it operates satisfactorily with the high values of pulse voltage and current encountered in radar. It is rugged both mechanically and electrically, and its long life is limited only by erosion of the fixed electrode. The gap must be shielded to prevent radio interference, and provision must be made for removing the corrosive gases generated by breakdown of the air in the spark. In air-borne systems, the gap is always enclosed in a pressurized container, and blocks of activated carbon are used to absorb the gases; in ground or shipboard applications, forced ventilation removes the gases. An erratic variation of ± 50 μs in the time of firing requires that self-synchronous timing systems be used with rotary spark gaps.

The action of the gap begins when a pin on the rotating disk approaches

a fixed electrode which is connected to the pulse-forming network. A pointed arm adjacent to the fixed electrode generates ions by corona discharge which assists the breakdown of the gap. As the space between electrodes decreases, the field strength increases until the gap fires and becomes practically a short circuit. The circuit remains effectively

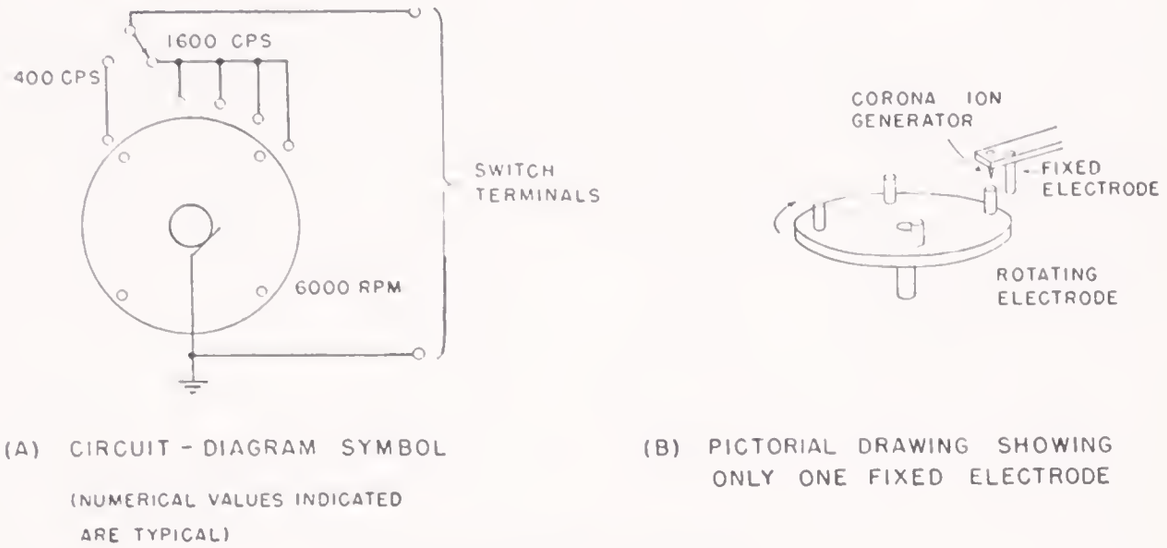


FIG. 8. Rotary spark gap.

closed during discharge of the pulse-forming network. Following the pulse, the voltage between electrodes is held to a value too low to sustain ionization, and after a deionization interval of a few tens of microseconds the circuit is open and the network begins to recharge.

If more than one repetition rate is desired, several fixed electrodes may be employed, as shown in Fig. 8*A*. The pulse rate is shifted by switching

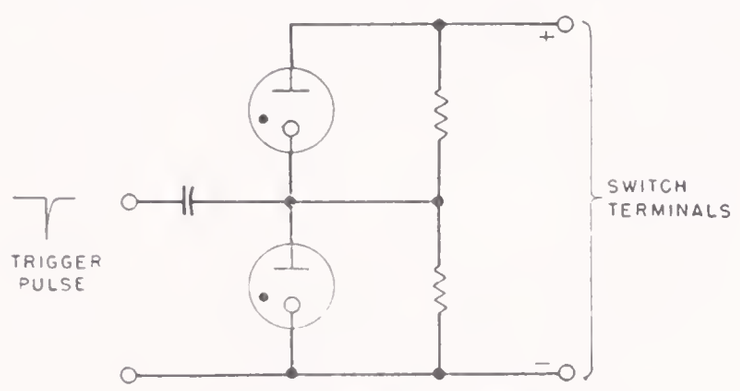


FIG. 9. Gas-diode switch.

some of the fixed electrodes in or out of the circuit. One rotary spark gap may be employed to discharge several parallel pulse-forming networks, each connected to its own load, and the gap may carry pulse currents in excess of 1,000 amp.

Gas Diodes. Two or more cold-cathode gas diodes, such as Type 1B22 tubes, may be connected in series to form a switching device, as shown for two tubes in Fig. 9. Equal resistances connected in parallel with the tubes divide the total voltage of the network equally between the tubes

so that the voltage across each is not enough to make it conduct. A trigger pulse increases the voltage across one tube and causes it to conduct. Several large diodes in series make a suitable switch for very high-power modulators. The life of some cold-cathode gas diodes is greatly decreased if the rated pulse duration or pulse current is exceeded even slightly.

In Fig. 9, the negative trigger pulse causes the upper tube to conduct. The voltage across this tube then falls to a very small value (approximately 60 volts), and nearly all the network voltage appears across the lower tube, causing it to conduct and close the circuit. The circuit remains closed until the tubes deionize, and then the charging cycle for the network begins.

The chief limitation on the use of gas diodes as a switch is the very large trigger pulse required to start conduction. Another difficulty is the uncertainty of the firing time of the second tube.

Argon Thyratrons. The permissible peak voltage of argon thyratrons, such as Type 2050, is about 1 kv. Therefore, these tubes are used only in driver circuits. The tube is normally held cutoff by a negative grid voltage and is made to conduct by a positive trigger pulse applied to the grid. Once fired, the tube continues to conduct until the current furnished by the pulse-forming network falls below the level required to sustain ionization. After a deionization interval in which the plate-to-cathode voltage must be kept low, the negative grid voltage regains control, and the network can recharge. Firing time of the thyratron is very precise.

Hydrogen Thyratrons. The hydrogen thyratron is a very effective switching device for use in a transformer-coupled line-pulsing modulator. The Type 4C35 thyratron can be used with a pulse-network voltage of more than 8 kv and can carry 90 amp with a drop across the tube of approximately 70 volts.

The average life of the hydrogen thyratron is about 900 hr at a pulse repetition rate of 2,000 per second. Since the life depends upon the total number of pulses, it is approximately inversely proportional to repetition rate. Tube life is unaffected by pulse duration in the range from 0.1 to 5 μs. The tube is fired by a 30- to 150-volt trigger pulse applied to the grid.

Saturable-core Inductors. An inductor may be designed so that its core saturates for coil currents much smaller than those that are commonly present. The incremental inductance is then high (for example, 1 h) for currents less than the saturation value, and low (for example, 10^{-4} h) for currents in the saturation range. Saturation of the core has an effect in the circuit comparable to firing a spark gap. Saturable-core inductors may be designed for any required values of voltage and current and are

used in line-pulsing modulators that drive magnetrons directly without matching transformers. Since an inductor is neither an electronic nor a mechanical device, it has a very long life. The shunt capacitance of the saturable-core inductor is high, the efficiency of the associated charging circuit is relatively low, and the pulse shapes obtained are poorer than those obtained using other switching devices.

TABLE 1. SUMMARY OF PROPERTIES OF SWITCHING DEVICES FOR USE WITH PULSE-FORMING NETWORKS

Property	*Rotary spark gap*	*Gas tubes*			*Saturable-core inductor*
		1*B*22 *gas diode*	2050 *argon thyratron*	4*C*35 *hydrogen thyratron*	
Firing precision	Jitter* ± 50 μs	Jitter* ± 6 μs	Good	Good	Good
Switch closed by	Mechanical motion	Trigger pulse	Trigger pulse	Trigger pulse	Saturation current
Switch opened by	Deionization after current from network has fallen below minimum to sustain ionization				Current falling below saturation point
Peak pulse current, amp	More than 100†	About 70	More than 20	About 90	More than 15
Peak voltage, kv	More than 30	3.6 per tube	About 1	More than 8	More than 20
Normal life, hr	Adjust every 100	50–500*	200–500	500–900	Several thousand
Other properties	Removal of corrosive gases and r-f shielding required Sensitive to atmospheric conditions	Short life on long pulses Very large trigger pulses required*	Used only in driver circuits	Short life with high pulse rate	Poor pulse shape* Large shunt capacitance Special charging circuit

* Major disadvantage or limitation.

† Most rotary spark gaps have a pulse-current rating slightly higher than 1000 amp.

B. LINE-PULSING MODULATORS

A block diagram of a line-pulsing modulator using an open-end type of pulse-forming line is shown in Fig. 10. Between pulses the switch is

open, and the power supply charges the network through the charging impedance and the primary winding of the pulse transformer. During a pulse, the switch is closed, and the network discharges through the switch and the primary winding of the pulse transformer and supplies energy to the transmitter. The apparent impedance of the transformer primary

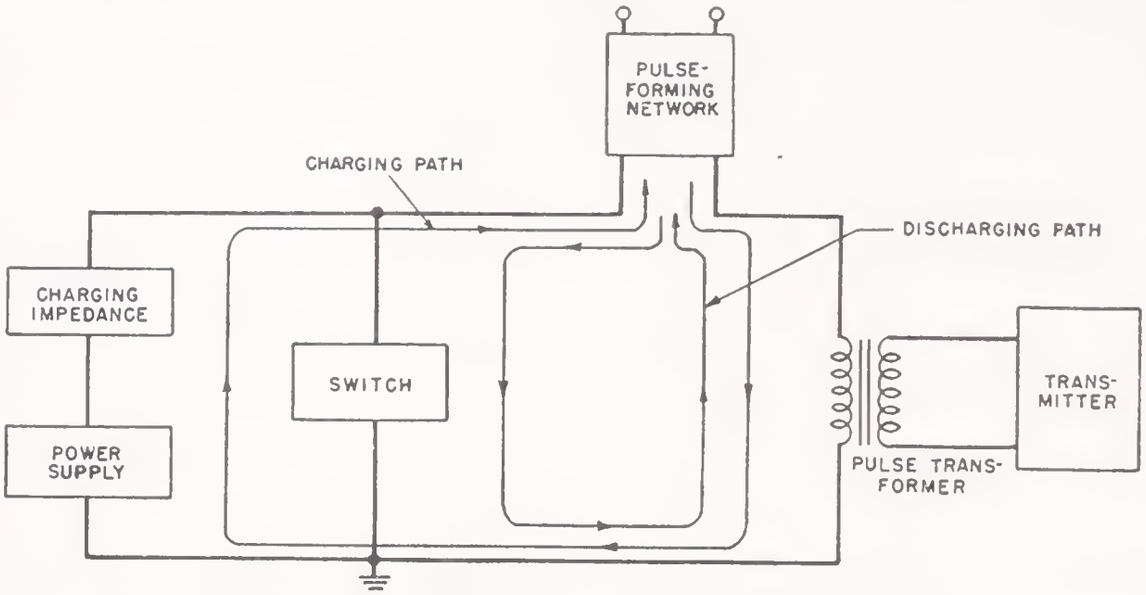


FIG. 10. Block diagram of a line-pulsing modulator.

equals the characteristic resistance of the network; therefore the form of the voltage pulse applied to the transmitter approaches a rectangle.

7. Charging Methods. The charging circuit of Fig. 10 may be represented as an a-c or d-c power supply in series with a charging impedance and the storage capacitance of the pulse-forming network, as shown in Fig. 11. The transformer primary winding is omitted from Fig. 11

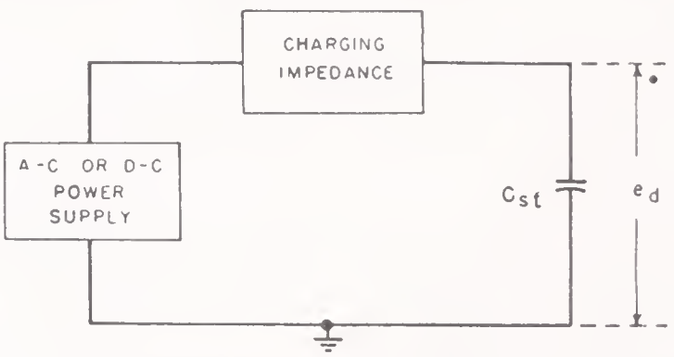


FIG. 11. Charging circuit for pulse-forming network.

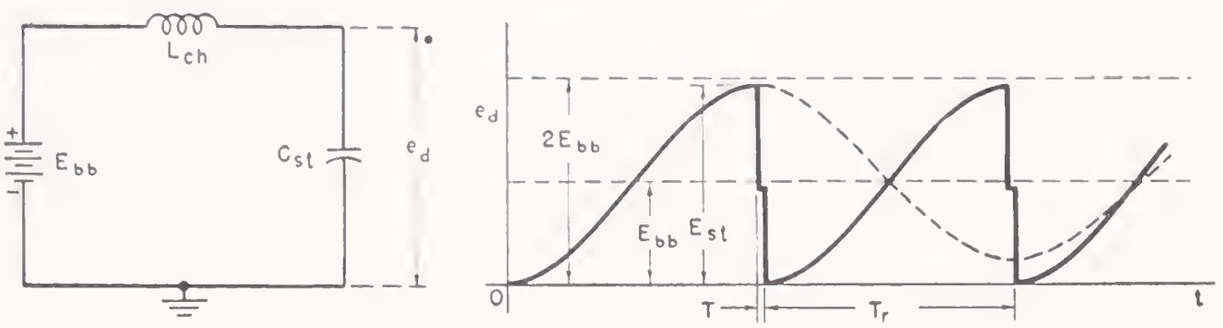
because its impedance is too low to affect the charging process. The pulse-forming network is represented by its storage capacitance because the charging process is so slow that, as explained in Art. 1, the line inductances may be considered short circuits. The value of the storage capacitance may be calculated in terms of T and R_c by means of Eq. 1, p. 164.

The obvious method of charging a pulse-forming network is d-c resistance charging, in which the power supply of Fig. 11 is a source of direct voltage and the charging impedance is a resistor. This method is

employed in driver stages where the power level is low, but it is never used in line-pulsing modulators because its efficiency is only 50 per cent. During the process of placing a charge Q_{st} on the storage capacitor C_{st}, a power supply of voltage E_{bb} supplies an energy $Q_{st}E_{bb}$ to resistor and capacitor together. The final capacitor voltage is E_{bb}, and the stored energy is therefore $\frac{1}{2}C_{st}E_{bb}^2$ or $\frac{1}{2}Q_{st}E_{bb}$. Thus half of the input energy is dissipated in the resistor.

Several more efficient charging methods, which are used in line-pulsing modulators, are discussed in the following paragraphs. An additional charging method, inductive-impulse charging, used when a saturable-core inductor is employed as a switch, is discussed in Art. 9.

Direct-current Resonance Charging. Figure 12*A* illustrates a d-c resonance-charging circuit which employs a d-c power supply E_{bb} in series



(A) EQUIVALENT CHARGING CIRCUIT

(B) WAVEFORMS OF e_d (DOTTED LINE APPLIES IF SWITCH IS NEVER CLOSED)

FIG. 12. D-c resonance charging.

with a charging inductance L_{ch}. Immediately following a pulse, the charging process begins with the capacitance C_{st} uncharged and no current in L_{ch}. Since the resistance of the series L-C circuit is low, the voltage across C_{st} oscillates sinusoidally (as shown by the dashed line in Fig. 12*B*) and, during the first cycle, reaches a maximum value E_{st} approximately equal to $1.9E_{bb}$, this value being slightly less than $2E_{bb}$ because of losses in the circuit elements. At the time of maximum voltage, the current in L_{ch} is zero. If the switch closes at this instant, a pulse of voltage $E_{st}/2$ is produced, and the charging process repeats as shown by the solid line in Fig. 12*B*. Successive charging processes are alike because the capacitor voltage and inductor current are each zero immediately following the pulse.

To obtain maximum stored voltage, E_{st}, the switch should be timed to fire at the peak of the voltage oscillation; that is, since the time duration of the pulse is negligible, the repetition period T_r should be half the period of the charging oscillation. Thus

$$T_r = \frac{2\pi\sqrt{L_{ch}C_{st}}}{2} = \pi\sqrt{L_{ch}C_{st}} \tag{3}$$

Moderate deviations of T_r from this value cause only a slight change in the value of E_{st}.

Advantages of d-c resonance charging are that the storage voltage is almost twice the power-supply voltage and the charging efficiency is approximately 95 per cent. A disadvantage is the relationship between repetition period, storage capacitance, and charging inductance that must be maintained.

Direct-current Resonance Charging with a Diode. To permit operating a radar set at more than one repetition rate, a charging diode may be connected in series with the charging inductance, as shown in Fig. 13*A*. The capacitor is thus prevented from discharging after the maximum voltage E_{st} is reached, except for a slight loss of charge due to leakage in

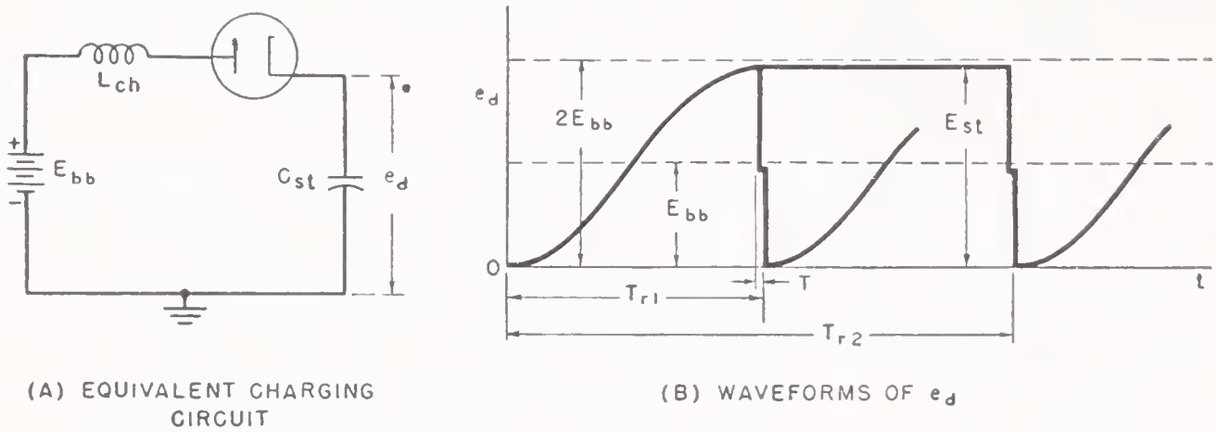


FIG. 13. D-c resonance charging with a diode.

the circuit elements. The charging process is identical with that of Fig. 12 except that the diode causes some additional loss, so that E_{st} is about 1.7 to 1.8 times E_{bb}.

Because the diode maintains the voltage at about $1.8E_{bb}$ for a considerable time, the interval between times of closing the switch can have any value greater than a half cycle of the charging oscillation, as indicated by T_{r1} and T_{r2} in Fig. 13*B*.

Normally, the charging circuit is designed so that the voltage e_d reaches its maximum value E_{st} in somewhat less time than the shortest repetition period $T_{r_{\min}}$ of the radar set. The charging inductance L_{ch} must not be larger than the value found from Eq. 3 with this minimum repetition period substituted for T_r. The charging inductance can be made considerably smaller than this value, but an excessively small charging inductance reduces the voltage E_{st} because the damping increases as L_{ch} is decreased.[1]

[1] If R_{ch} represents the total series resistance that causes damping in the charging circuit, then a sine wave must be multiplied by the factor $\epsilon^{-tR_{ch}/(2L_{ch})}$ to represent the damped oscillations of e_d. At the peak of the oscillation, t is approximately $\pi\sqrt{L_{ch}C_{st}}$, so that the multiplying factor for E_{st} is $\epsilon^{-(\pi R_{ch}/2)\sqrt{C_{st}/L_{ch}}}$. Therefore, the smaller L_{ch} is made, the smaller the peak value to which e_d rises. The reduction of E_{st} is not appreciable, however, unless L_{ch} is made considerably less than the value specified by Eq. 3.

The efficiency with a diode is slightly lower than without a diode, but still it is nearly 90 per cent. Both types of d-c resonance charging are used in radar systems.

In a line-pulsing modulator employing a hydrogen thyratron and d-c resonance charging with a diode, a special difficulty arises because the thyratron conducts in only one direction. In Fig. 14 ignore for the moment the by-pass diode, and suppose the pulse-forming network (represented by C_{st}) to be charged to the voltage E_{st} and then discharged through the thyratron and transformer. A slight mismatch during discharge may cause a small voltage of polarity opposite to E_{st} to appear across the network after the pulse. This small reversed voltage is in series with the battery at the start of the ensuing charging period. Hence the total voltage available to produce the charging oscillation is increased,

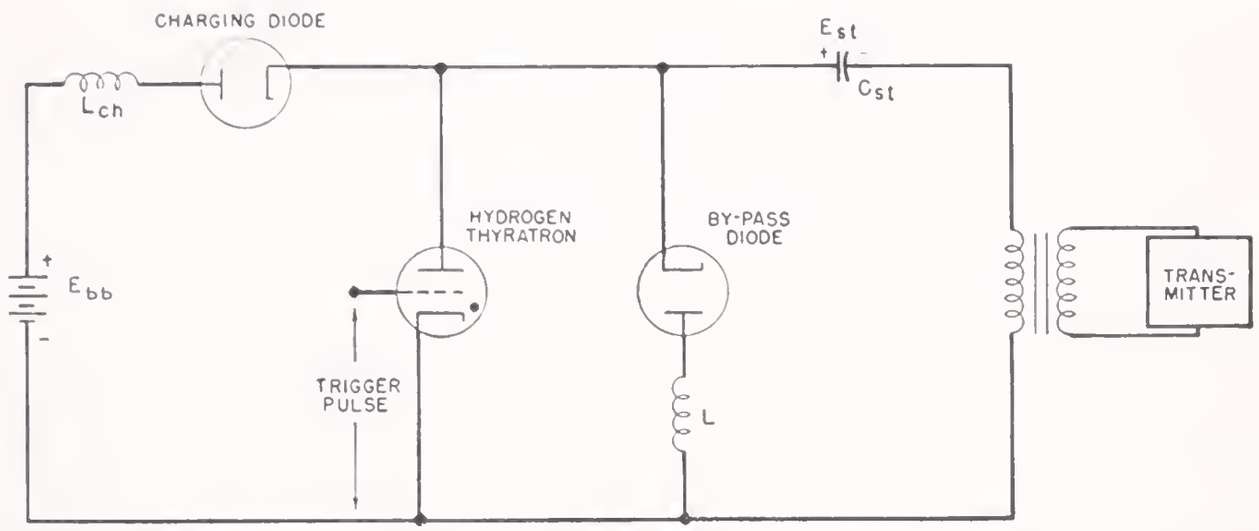


FIG. 14. Discharge circuit for modulator using a hydrogen thyratron and d-c resonance charging with a diode.

and the voltage stored on the network for the next pulse is greater than the first value of E_{st}. The next discharge leaves a greater magnitude negative charge on the network for two reasons: (1) It occurs from an increased voltage level and therefore would produce a greater final voltage even if the magnetron were a linear resistance. (2) Because of the nonlinear nature of the magnetron resistance (Art. 5), the increased voltage increases the degree of mismatch between the network and its load. After several cycles of operation with successively greater values of stored and residual voltages, the thyratron voltage may become greater than its maximum allowable value.

The trouble is overcome by connecting a by-pass diode and a series inductance L in parallel with the thyratron, as shown in the figure. At the end of the pulse, the reversed charge on the network escapes through the diode. The series combination of the pulse-transformer inductance and the inductance L in resonance with C_{st} causes an oscillation that reverses the voltage on the network. This new small voltage has a

polarity in the normal direction and tends to reduce slightly the maximum voltage on the network during the following charging period.

Alternating-current Diode Charging. The pulse-forming network may be charged by an a-c power supply in series with a diode (see Fig. 15*A*). At the a-c line frequency, the resistances of the load and the diode are negligible compared with the reactance of the storage capacitor.[1] To a good approximation, therefore, the alternating voltage is applied directly across capacitance C_{st}, and e_d follows the applied voltage from zero up to the peak value of the supply voltage, as shown in Fig. 15*B*. The diode prevents the capacitor from discharging as the power-supply voltage drops. Therefore E_{st} is approximately E_m.

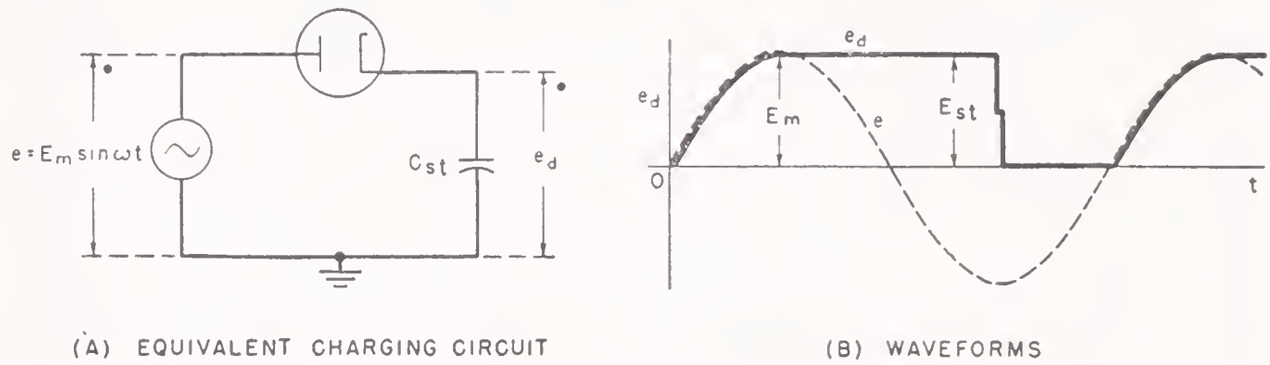


FIG. 15. A-c diode charging.

If the switching device were to discharge the network at any instant when the power-supply voltage is in the direction to cause the diode to conduct, the diode and switching device would form a short circuit across the power supply. Hence the switch must be synchronized to produce pulses only during half cycles of the a-c power supply in which the diode cannot conduct. The pulse repetition period is usually equal to the period of the a-c supply, though it may be any multiple of this period. The efficiency of a-c diode charging is usually higher than 75 per cent.

Alternating-current Resonance Charging. In another method of charging a pulse-forming network from an a-c source, the network is connected in series with an inductance L_{ch} and an alternating voltage of peak value E_m, as shown in Fig. 16*A*. Inductance L_{ch} is chosen so that the resonant frequency of the charging circuit is equal to the frequency of the a-c supply voltage.

[1] The reactance of the storage capacitor is $X_C = 1/2\pi f_{ac} C_{st}$, where f_{ac} is the frequency of the a-c supply. The period of the a-c supply is $T_{ac} = 1/f_{ac}$, and C_{st} (according to Eq. 1) is $T/2R_c$. Thus

$$X_C = \frac{R_c}{\pi}\frac{T_{ac}}{T}$$

The value of R_c is ordinarily about 50 ohms, and the ratio T_{ac}/T may vary from several thousand to tens of thousands. Thus X_C is of the order of 50 K at least. The resistance of a diode is only a few hundred ohms.

Waveforms of the supply voltage e and the network voltage e_d are shown in Fig. 16*B*. In this figure the supply voltage is assumed to be applied at $t = 0$, with the initial current in L_{ch} and voltage across C_{st} each zero, and the switching device is assumed to remain open so that no pulses are produced. The final amplitude to which the oscillations of e_d increase is much greater than the amplitude of the supply voltage because e_d is the voltage across the capacitor in a low-loss series resonant circuit. A number of cycles are required for e_d to reach its final high amplitude, however, because the energy associated with the oscillation is

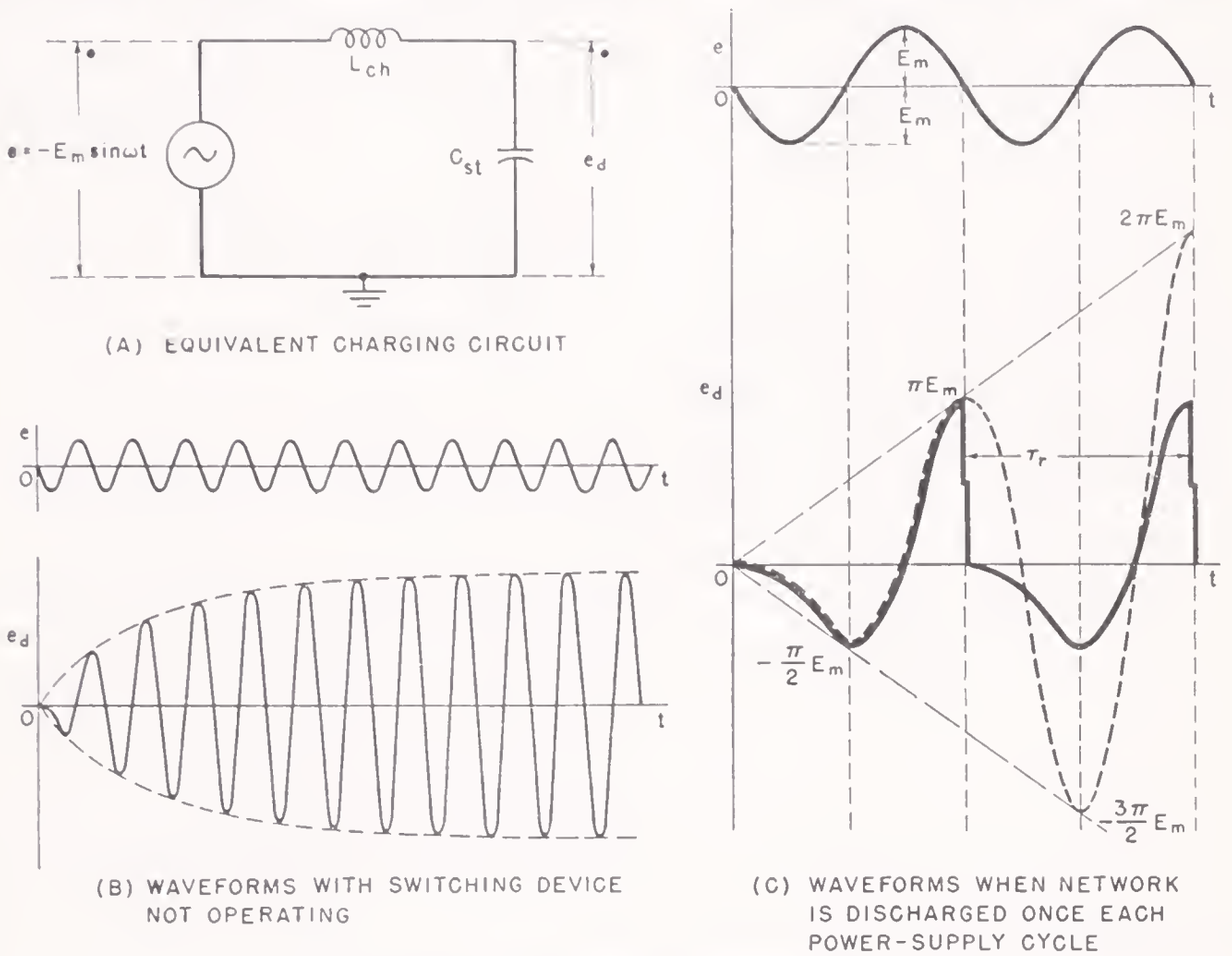


FIG. 16. A-c resonance charging.

supplied from the power source a little each cycle. It can be shown that, if damping is neglected, energy is absorbed from the power source at a rate which makes the successive peak voltages reached by e_d equal to $-1\frac{1}{2}\pi E_m$, πe_m, $-3_2\pi E_m$, $2\pi E_m$, and so on. Damping in the circuit has little effect on the first few peaks, but it makes the peaks approach a limiting value after several cycles instead of increasing indefinitely. The peaks occur 90 deg later than the corresponding peaks of the supply voltage, beginning with the first cycle.

In Fig. 16*C*, the first two cycles of e_d are shown as a dashed curve. In addition, a solid curve shows the waveform of e_d obtained when the pulse-forming network is discharged at a repetition frequency equal to the line frequency. The first cycle follows the dotted curve, and a voltage πE_m is

reached. The current in L_{ch} is zero when the voltage reaches its peak value because of the 90-deg difference of phase between capacitor current and voltage. At this instant the network is discharged. Thus the next cycle is started with both the inductance current and the capacitor voltage zero, and the second cycle is identical with the first.

An alternate charging circuit, which uses half-cycle instead of full-cycle charging intervals, is shown in Fig. 17*A*. Here the power-supply voltage e is connected across the pulse-forming network. The leakage inductance of the power-supply transformer (indicated as a broken-line coil in Fig. 17*A*) is the charging inductance. The network is charged first in one direction and then in the other, as shown in Fig. 17*B*; but the rectifying spark gap, which is driven at a rate of one-half revolution per cycle of

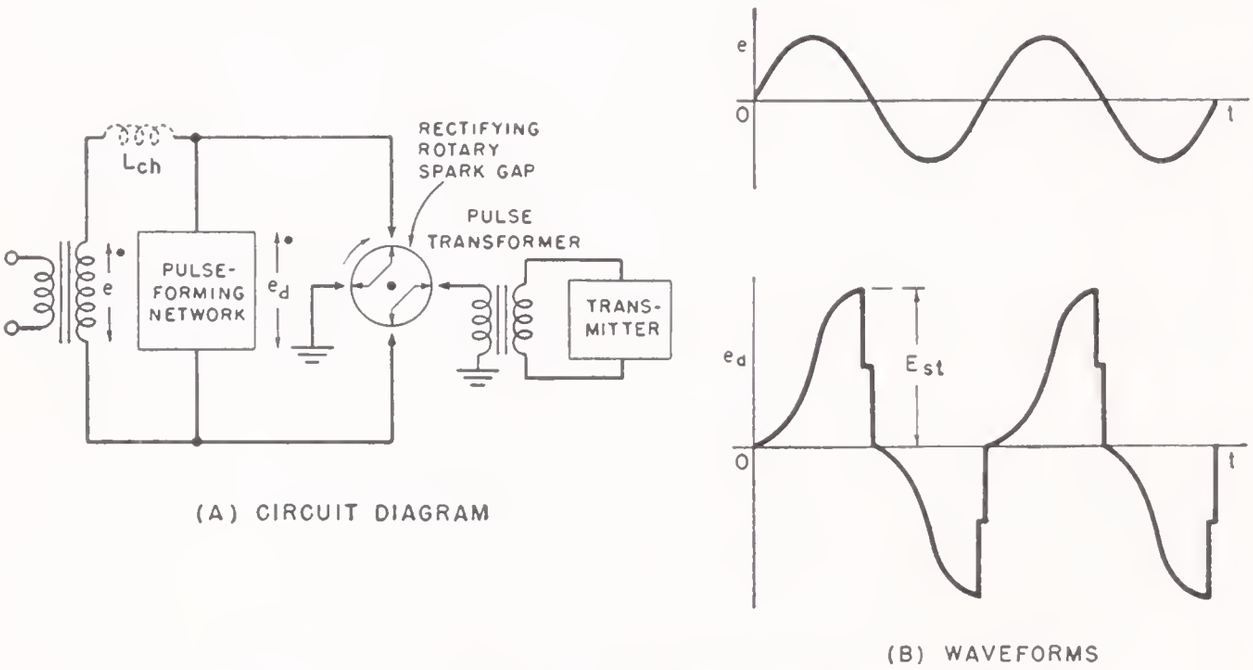


FIG. 17. A-c resonance charging with rectifying spark gap.

the a-c supply, delivers all pulse voltages to the pulse transformer with the same polarity.

If half-cycle charging is used with the rectifying spark gap, a repetition rate of twice the power-supply frequency is obtained. With half-cycle charging, damping reduces the value of E_{st} from $E_m\pi/2 = 1.59E_m$ to about $1.5E_m$. When 1-cycle charging is used, as in Fig. 16*C*, one pulse for each power-supply cycle is obtained. Instead of πE_m, the stored voltage E_{st} is about $2.8E_m$ because of damping.

A greater gain of voltage is obtained if more charging cycles are allowed between pulses, but the pulses are more widely spaced. Thus, with 2-cycle charging, one pulse is obtained for each two power-supply cycles, and E_{st} is nearly $2\pi E_m = 6.28E_m$. Losses reduce E_{st} to about $5E_m$.

8. Rotary-spark-gap Modulator. In Fig. 18 is shown the circuit diagram of a typical line-pulsing modulator in which a rotary spark gap is used to discharge the pulse-forming network. The simplified diagrams

of Fig. 19 indicate the circuits for charging the network slowly between pulses and for discharging it suddenly to form a pulse.

Description of Circuit. The voltage-doubling rectifier provides a nearly constant direct voltage of 4.5 kv for use in charging the network. Because d-c resonance charging with a diode is used, the pulse network is charged to about $1.8 \times 4.5 = 8.1$ kv. The time required by the charging process depends upon the charging inductance $L_{ch} = 15$ h and upon the storage capacitance of the pulse-forming network. According to Eq. 1, the storage capacitance is

$$C_{st} = \frac{T}{2R_c} = \frac{1}{2 \times 50} = 0.01\ \mu\text{f}$$

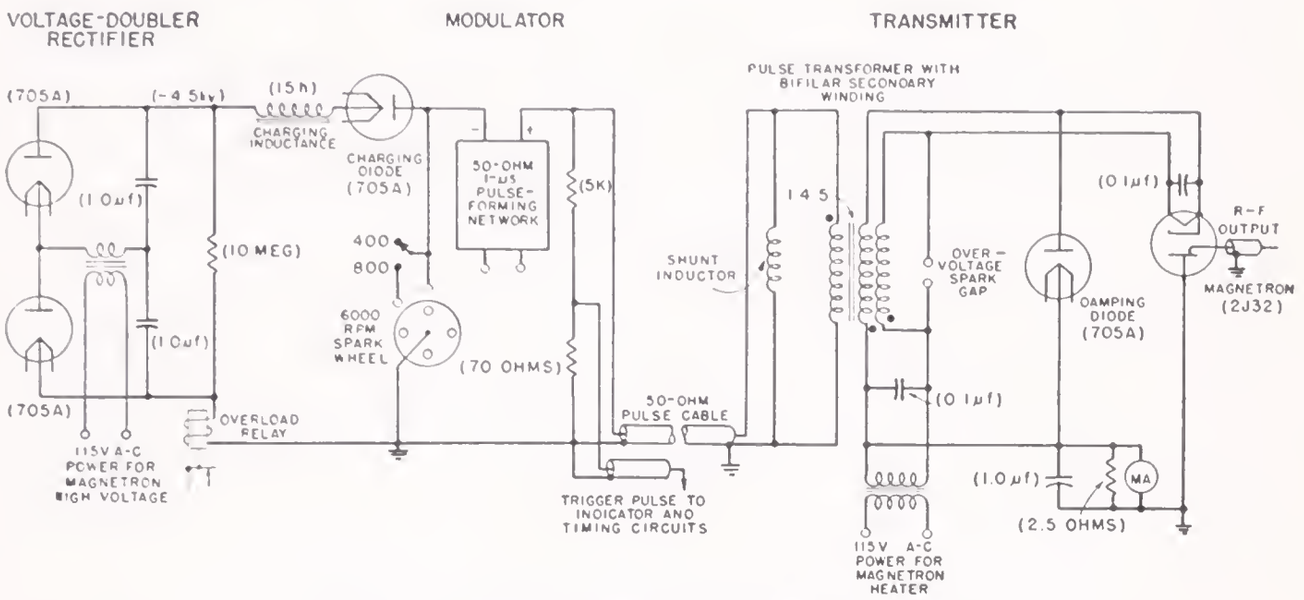


FIG. 18. Typical rotary-spark-gap modulator.

The minimum time taken to charge the network is a half period of the charging oscillation, or

$$\text{Minimum charging time} = \pi \sqrt{L_{ch}C_{st}} = \pi \sqrt{15 \times 10^4} = 1{,}220\ \mu\text{s}$$

Therefore the repetition frequency may not be higher than

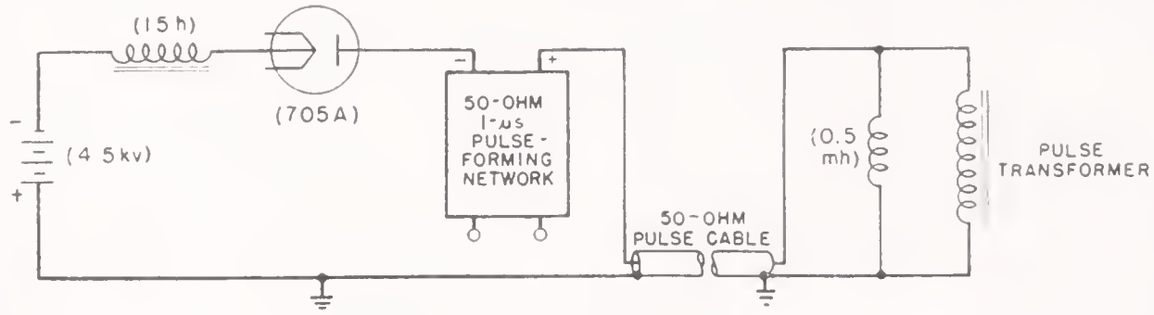
$$\frac{10^6}{1{,}220} = 820 \text{ cps}$$

Because of the diode, lower repetition frequencies are allowable.

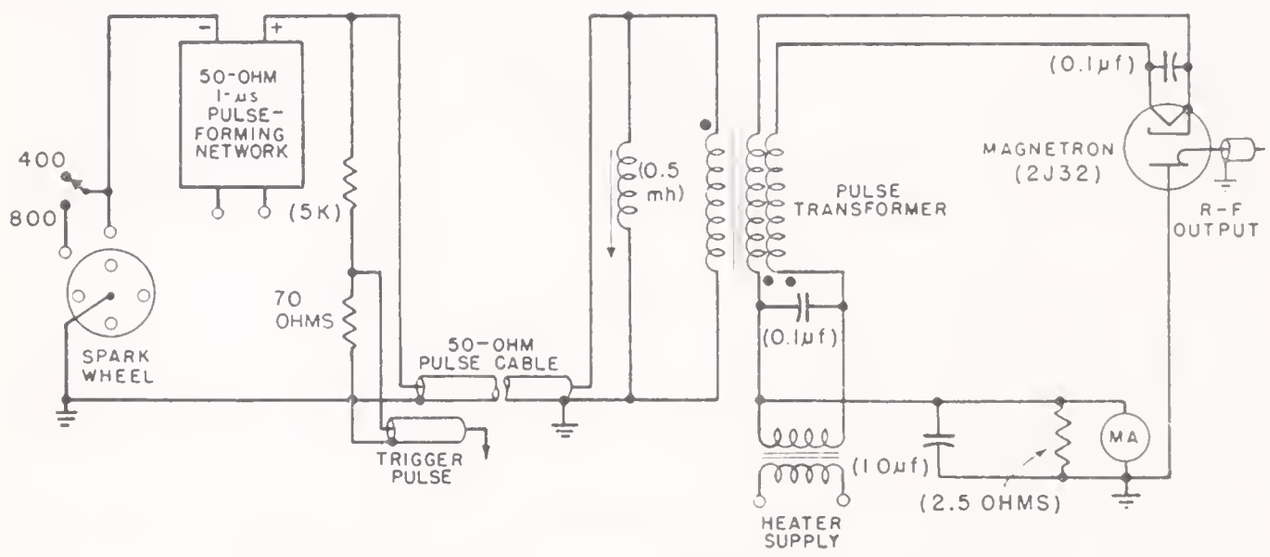
Since the spark wheel of the rotary spark gap turns at 6,000 rpm (100 rps) and has four electrodes, a repetition frequency of either 400 or 800 cps may be obtained—400 cps if one fixed electrode is used, 800 cps if both are used. Discharge of the pulse-forming network provides a 1-μs pulse. Sometimes two pulse-forming networks are used with a switching arrangement to allow a choice of pulse duration. For example, a 1-μs pulse might be used for normal radar operation, and a 2¼-μs pulse to trigger radar beacons for navigation purposes (see Art. 8, Chap. I).

The amplitude of the voltage pulse applied to the 50-ohm pulse cable is about 4.05 kv, that is, half the voltage to which the pulse network is charged. Frequently the transmitter is located close to the antenna to avoid a long r-f line, and the modulator is located at some distance from the antenna for convenience. A relatively long pulse cable is then required.

In parallel with the input to the pulse cable is a voltage divider that supplies a very small portion (about 55 volts) of the modulator pulse as a timing pulse for other parts of the system. Self-synchronous operation



(A) CHARGING CIRCUIT



(B) DISCHARGING CIRCUIT

FIG. 19. Equivalent charging and discharging circuits for modulator of Fig. 18.

is required because the transmitted pulse is timed by the rotation of the spark wheel, and the instant at which the gap fires is somewhat erratic.

The 4.05-kv pulse from the network and cable is increased in voltage by the pulse transformer, and a pulse of $4.5 \times 4.05 = 18.2$ kv is applied to the magnetron. Because the winding directions of the pulse-transformer coils are as indicated by the polarity dots, a negative pulse is applied to the cathode of the magnetron. The transformer serves two purposes: (1) It allows the pulse network to operate at a considerably lower voltage than the magnetron, thereby reducing insulation difficulties in most of the circuit. (2) It matches the apparent resistance of the magnetron load (about 800 to 1,000 ohms) to the 50-ohm pulse network and pulse

cable. Pulse cables are frequently long enough to have a delay time that is of the same order of magnitude as the pulse duration. Therefore, the characteristic resistance of the pulse network must be made equal to the low resistance of the cable, and the transformer is required to terminate the cable properly.

The secondary winding of the pulse transformer is *bifilar;* that is, it is made by laying two insulated conductors side by side so that exactly the same voltage is induced in each. By means of the bifilar winding, an alternating voltage is applied from a 6.3-volt filament transformer to the magnetron heater, and high-voltage pulses are applied between plate and cathode. If a single-wire pulse-transformer secondary winding were used, a filament transformer with high-voltage insulation between windings would be required. By-pass capacitors are sometimes placed at the ends of the bifilar winding, as in Fig. 18, to allow the pulse current to divide between the two parts of the bifilar winding without affecting the heater circuit.

An overvoltage spark gap is connected across the secondary terminals of the pulse transformer, and a d-c milliammeter is provided to indicate the average magnetron current. The 2.5-ohm shunt resistor is adjusted for correct reading of the milliammeter. Because of this resistor, the meter may be switched into another circuit without affecting the operation of the modulator. The 1.0-μf capacitor by-passes the a-c components of the pulse current, so that the meter carries a nearly steady current.

Magnetron Voltage after the Pulse. The shunt inductor and damping diode in Fig. 18 are needed to prevent the magnetron from producing low-power r-f oscillations and noise during a period of several mirco-seconds following the main pulse. The waveform of the magnetron cathode voltage obtained when the inductor and diode are omitted is shown in Fig. 20*A*. A rather complex voltage variation occurs after the main pulse, and a number of the negative peaks in it are of sufficient amplitude to cause the magnetron to oscillate. Thus false signals that may obscure real echoes are produced.

The high-frequency equivalent circuit for the pulse transformer drawn in Fig. 21 shows that the leakage inductance L_l and the parallel combination of transformer output capacitance C_o,* wiring capacitance C_w, and magnetron capacitance C_{mag} form an oscillatory circuit damped by the magnetron plate resistance r_p. The leading edge of the pulse starts oscillations that are rapidly damped by the low resistance of the conducting magnetron. The trailing edge starts oscillations that persist for a long time because the magnetron resistance is high after the pulse and

* Because of the high ratio of the transformer, C_o is very nearly the sum of the secondary capacitance and the interwinding capacitance.

provides little damping. These oscillations cause the largest part of the voltage variations after the pulse in Fig. 20*A*.

Reflections occur at the end of the pulse cable when the pulse begins and ends, even though the transformer is adjusted for a perfect match

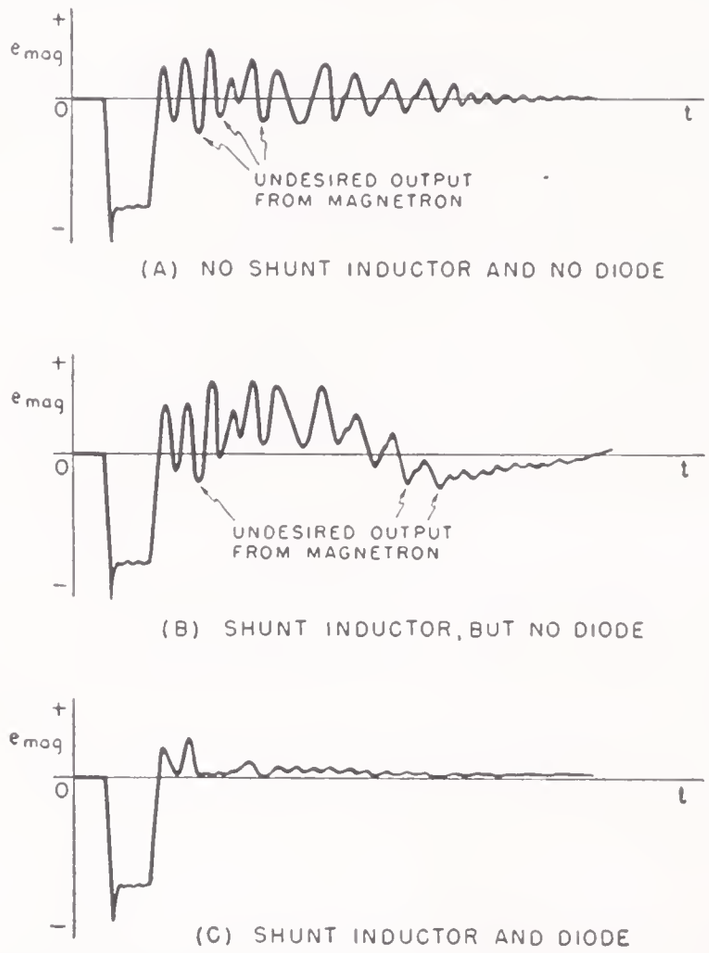


Fig. 20. Elimination of undesired magnetron output by means of inductor and diode.

during the main part of the pulse. Because of the capacitances in Fig. 21 and the nonlinear character of the magnetron plate resistance, the decrease of transformer-primary voltage at the end of the pulse is accompanied by a more than proportional decrease of current. Therefore a

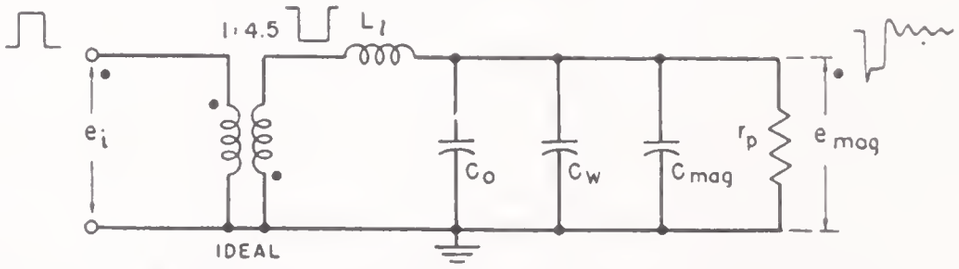


Fig. 21. High-frequency equivalent circuit for pulse transformer.

very brief positive voltage pulse is reflected when the main pulse ends as if the terminating resistance were increased. The positive pulse travels from the transformer to the pulse-forming network. Because sufficient time for deionization has not elapsed, the spark gap is conducting. Therefore the pulse enters the pulse-forming network, is reflected without

reversal of polarity at the open end of the network, and returns to the pulse-transformer primary. A negative pulse from the transformer secondary is applied to the magnetron cathode delayed by 1 μs plus the two-way travel time in the pulse cable. This pulse appears superimposed on the high-frequency oscillation in Fig. 20*A*. Additional negative pulses sometimes result from successive reflections.

To eliminate the undesired magnetron output, a positive component of magnetron voltage following the pulse is added by means of the shunt inductor. The effect of the inductor is shown in Fig. 20*B*, which shows a voltage like that of Fig. 20*A* but with an added low-frequency oscillation. During the pulse, current builds up in the inductor in the direction of the arrow in Fig. 19*B*. After the pulse, this current continues through the conducting spark gap into the pulse-forming network, producing a reversed voltage on the pulse-forming network and a positive voltage at the magnetron cathode. Thus an oscillation having a frequency determined by the storage capacitance of the network and the shunt inductor is started.

Use of the shunt inductor reduces the magnetron output immediately after the pulse but causes an output at later times during the negative half cycles of the low-frequency oscillation. The damping diode conducts when the magnetron voltage is positive and therefore changes the low-frequency oscillation of Fig. 20*B* to the overdamped *R-L-C* transient of Fig. 20*C*. In addition, the conducting diode provides considerable damping of the high-frequency oscillation and considerably reduces the amplitude of the pulses caused by reflections. As shown in Fig. 20*C*, the combination of inductor and diode eliminates all negative magnetron voltages after the pulse.

Other methods of eliminating the undesired negative voltages are sometimes used. Instead of a shunt inductor, a pulse transformer with low magnetizing inductance may be used. The magnetizing current of the transformer then serves the same purpose as the current in the inductor. Instead of a damping diode, a nonlinear resistor (such as Thyrite) in series with the spark wheel may be used. The resistance of such a circuit element is low during the main pulse when the current through it is high; therefore the loss of pulse power is not excessive. After the pulse, the current is low and the resistance of Thyrite is high; therefore it damps oscillations of the series circuit in which it is placed. Another method of providing damping is use of a transformer core having thick laminations so that the eddy-current losses are large. In some modulators a parallel combination of resistance and capacitance is placed in series with the magnetron. The voltage build-up on the capacitor during the pulse places a positive bias on the magnetron cathode for a short period after the pulse and thus helps to prevent the spurious output.

Operating Conditions during Pulse. At the beginning of the pulse the voltage applied to the pulse-transformer primary winding is 4.05 kv. Because the transformer matches the apparent resistance of the magnetron to the pulse cable, the primary winding presents a 50-ohm resistance to the cable, and the initial primary current is 4,050/50 = 81 amp. The initial magnetron current is therefore 81/4.5 = 18 amp, and the initial magnetron voltage is 4.05 × 4.5 = 18.2 kv. These values locate the operating point on the plate characteristics of the Type 2J32 magnetron (Chap. XI, p. 795). According to these curves, a flux density of 1,870 gausses is needed to permit operation at this point and to ensure a correct match for the pulse cable and network. The power output, read from the curves, is about 170 kw. The input power is 18.2 × 18 = 328 kw, so that the efficiency is $\frac{170}{328} \times 100 = 52$ per cent. This value checks the efficiency read from the curves.

Operating conditions change somewhat during the pulse because of the build-up of current in the shunt inductor. The amount of current built up can be calculated easily because the voltage across the inductor is very nearly constant. If e and i represent the inductor current and voltage, then during the pulse $e = L\Delta i/\Delta t$, or

$$4.05 \times 10^3 = 0.5 \times 10^{-3} \frac{\Delta i}{10^{-6}}$$

and $\Delta i = 8.1$ amp. Since the inductor current is zero at the beginning of the pulse, 8.1 amp is its value at the end of the pulse.

The increase of inductor current tends to decrease the voltage obtained from the pulse-forming network and consequently to decrease the magnetron current. Because of the shape of the magnetron characteristics, only a very small decrease of voltage is needed to cause a large decrease of magnetron current. Therefore, during the pulse, magnetron current decreases almost enough to compensate for inductor current, and little change occurs in either the magnetron voltage or the total current output of the pulse-forming network. The magnetron current decreases by nearly 8.1/4.5 = 1.8 amp, and the final operating point is located on the 1,870-gauss line at a current of 18 − 1.8 = 16.2 amp. Note that the final voltage is not appreciably changed from 18.2 kv. The power output at the end of the pulse is 155 kw.

The average power output during the pulse (generally called the peak power of the radar system) is

$$\frac{170 + 155}{2} = 163 \text{ kw}$$

The average plate current during the pulse is

$$\frac{18 + 16.2}{2} = 17.1 \text{ amp}$$

Since the pulse duration is 1 μs and the repetition period may be either 2,500 or 1,250 μs, the reading of the magnetron-current milliammeter is either 17,100/2,500 = 6.85 ma or 17,100/1,250 = 13.7 ma, the higher reading being obtained with the higher repetition frequency.

Deionization of Spark Gap. If the pulse-forming network is open-circuited during operation of the modulator (as may occur accidentally during switch-over from one network to the other if two networks are used), an arc encircles the entire spark wheel and the overload relay (see Fig. 18) interrupts power to the circuit. With the pulse-forming network out of the circuit, an arc once started continues uninterrupted as the spark wheel turns and short-circuits the power supply.

In normal operation the pulse-forming network is completely discharged during each pulse. Since the network charges slowly through the large charging inductance, the voltage across the spark gap is maintained at a low value for several hundred microseconds after each pulse. In this way time is provided for ions formed in the gap during the pulse to disappear so that the air between the electrodes may regain its insulating property. When the network is open-circuited, however, no means of maintaining a low spark-gap voltage is present; the mechanical separation of electrodes produced by rotation of the spark wheel cannot stop conduction in the ionized air; and a continuous arc is obtained. In every modulator that uses a gaseous-conduction switching device, the initial charging rate for the pulse-forming network must be low enough to permit deionization.

9. Modulator with Saturable-core Inductor and Inductive-impulse Charging. The circuit diagram of a line-pulsing modulator employing inductive-impulse charging and using a saturable-core inductor as a switching device is shown in Fig. 22. Operation of the circuit begins approximately 200 μs before a pulse is to be transmitted, at which time the normally off pulser tube (V_1 in Fig. 22) is turned on by a positive gate pulse from a timing multivibrator.

The waveform of the input pulse e_i is shown in Fig. 23*C*. Before the application of this pulse, capacitors C_F and C_{st} had each charged to 1,200 volts with the polarities indicated in Fig. 22. Therefore, as soon as V_1 begins to conduct, plate current for it is supplied through three paths: (1) The 1,200-volt power supply causes current to build up in the 200-mh charging inductor L_{ch} (see waveform of i_{ch}). (2) The 1,200-volt charge on capacitor C_F causes current to build up in the saturable-core reactor L_{sc} (see waveform of i_{sc}). At first i_{sc} builds up very slowly because L_{sc} is operating in the unsaturated region where it has an inductance of about 1 h. (3) The 1,200-volt charge on C_{st} causes current to flow through V_1 and the magnetron and bifilar inductor (see small pulse in waveform of i_{mag}). This current pulse is of short duration because C_{st} is small and is

quickly discharged. Nevertheless it could cause an undesired r-f output pulse. Therefore the 300-μh inductor L_r is included in the circuit to limit the rate at which C_{st} may discharge and to prevent appreciable magnetron output.

When current in L_{sc} increases to the saturation point (about 20 ma), the inductance of L_{sc} decreases to about 100 μh, and therefore a sudden increase of current occurs, as shown in Fig. 23*C*. The large value reached by i_{sc} quickly discharges C_F, and i_{sc} is then limited to the current that can flow through the 18-K resistor R_F.

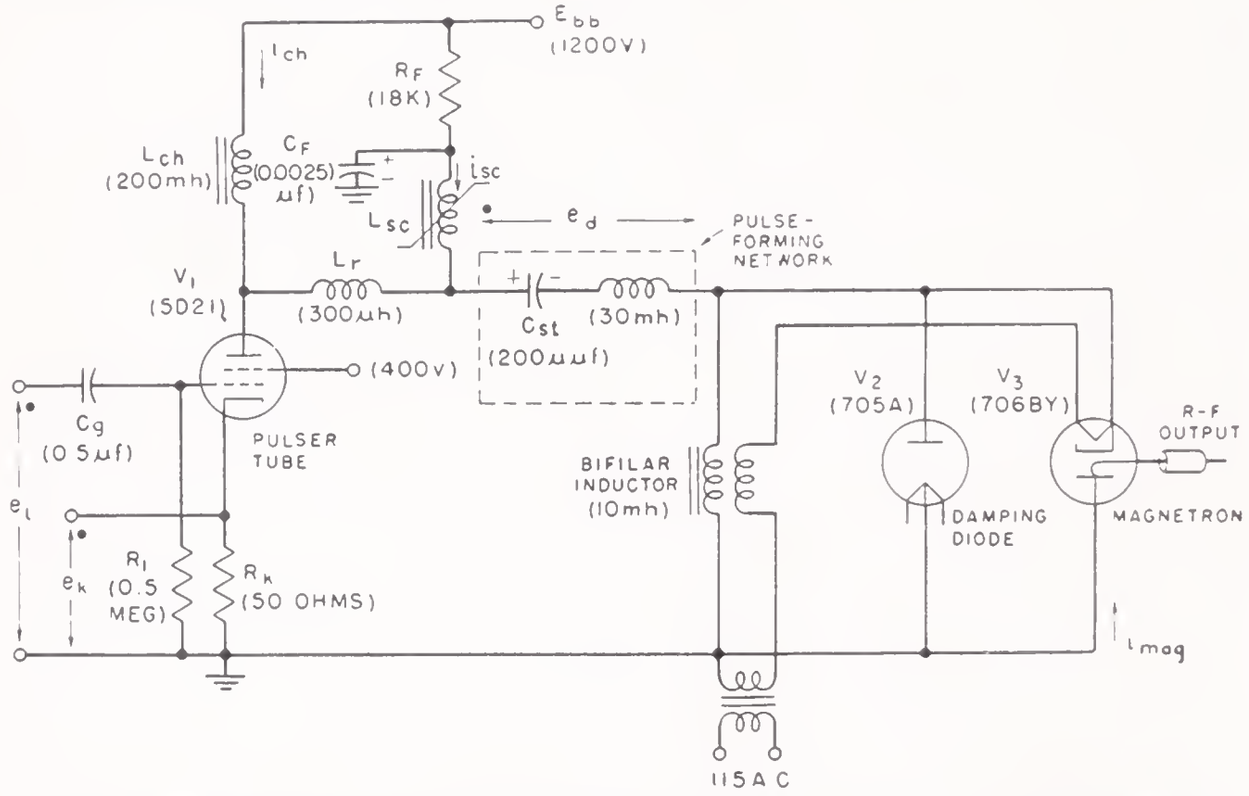


FIG. 22. Modulator using saturable-core inductor and inductive-impulse charging.

The 180-μs period during which tube V_1 is conducting is used to build up a current in the charging inductance L_{ch}, in preparation for the charging of C_{st}. To end the preparation period, the voltage e_k at the pulser-tube cathode is fed back to the multivibrator that supplies the input voltage. When i_{ch} reaches 1.2 amp, e_k becomes large enough to trigger the multivibrator, thus causing V_1 to be cut off very suddenly.

The current in L_{ch} cannot stop instantly when tube V_1 is cut off. Therefore i_{ch} flows into the pulse-forming network and charges C_{st} to a high voltage (see charging circuit of Fig. 23*A* and waveform of e_d in Fig. 23*C*). The maximum voltage to which C_{st} is charged is about 18 kv, the exact value being difficult to calculate because of the effects of damping and stray capacitances. Note that a voltage many times greater than the power-supply voltage is obtained by means of the *inductive impulse*. The time required to charge C_{st} is approximately one-quarter cycle of the oscillation frequency of the charging circuit, or about 10 μs for the circuit of Fig. 23*A*.

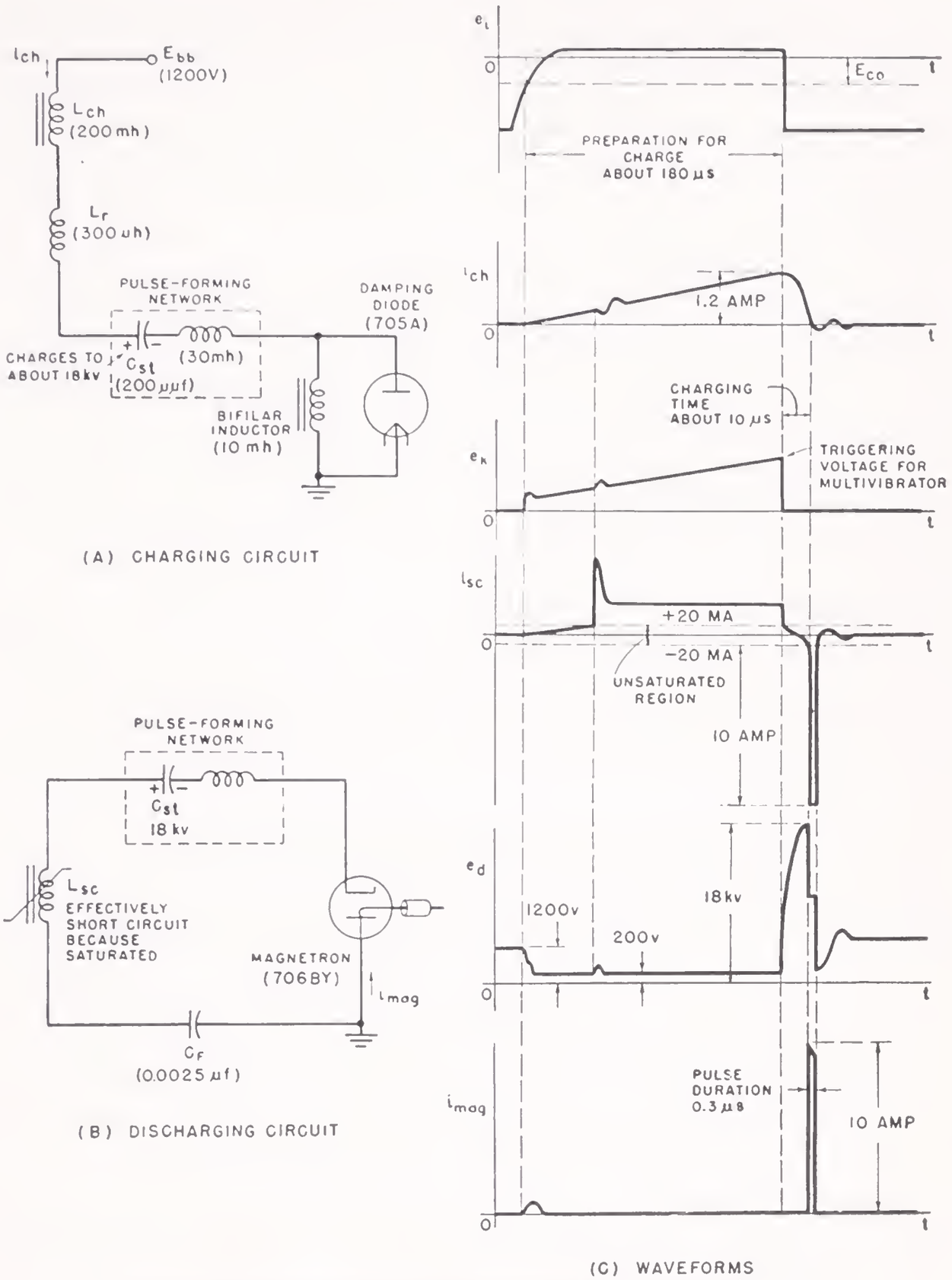


FIG. 23. Equivalent circuits and waveforms for modulator of Fig. 22.

When V_1 is cut off, current in the saturable-core inductor decreases very quickly to the limit of the unsaturated region. During the remainder of the charging period, the current in L_{sc} changes very slowly from about +20 ma to approximately −20 ma because the inductance of L_{sc} is high. Hence, the saturable-core inductor conducts a negligible current during the charging period and may be considered as an open switch.

The inductance of L_{ch} is adjusted so that the voltage e_d reaches its maximum value at the same instant that the current i_{sc} reaches the negative-current saturation limit. Further change of i_{sc} is extremely rapid and allows the pulse-forming network to discharge through the circuit of Fig. 23*B* almost as though the saturable-core inductor were a short circuit. The discharge path includes the magnetron, and therefore an output pulse is produced.

The 10-mh bifilar inductor in Fig. 22 is used to supply heater voltage to the magnetron in the same way that a bifilar secondary winding on a pulse transformer is used in the rotary spark-gap modulator of Fig. 18, p. 177. The bifilar coil also serves with the damping diode to prevent the magnetron from producing unwanted pulses after the main pulse (see Art. 8). The damping diode is also part of the charging path for the pulse network. Sometimes a small positive bias voltage (about 300 volts) is placed on the magnetron cathode to help in preventing undesired magnetron output.

This modulator circuit has the advantage of being simple, containing no moving parts, and requiring a relatively low-voltage power supply. It has the disadvantages, however, of low efficiency and poor pulse shape. The poor pulse shape is a result of the imperfections of the saturable-core inductor as a switching device and of the simple type of pulse-forming network used.

C. DRIVER-POWER-AMPLIFIER MODULATORS

Modulators of the driver-power-amplifier type consist of a driver circuit that forms pulses and a power amplifier that amplifies the pulses

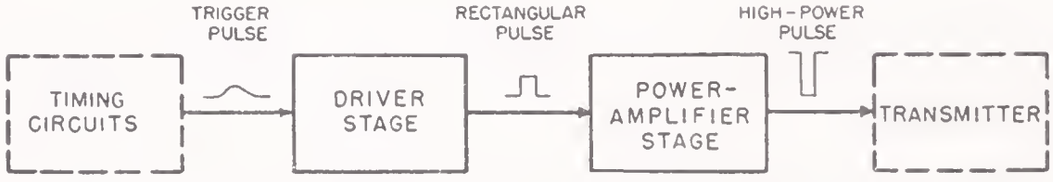


FIG. 24. Block diagram of driver-power-amplifier modulator.

and applies them to the transmitter. A driver-power-amplifier system is illustrated in block-diagram form in Fig. 24.

The power-amplifier stage is a high-power video amplifier. Special tubes with high current and voltage ratings are required, and often two or more tubes are connected in parallel. Because of the large powers involved, high efficiency of the circuit and effective utilization of the tubes are important. The power-amplifier tubes are always cutoff between pulses to avoid unnecessary loss of power, and during the pulse the grid voltage is driven far positive so that the amplifier tubes can carry large currents with low plate-to-cathode voltage.

The driver stage of the modulator must also use large tubes, because it must furnish grid power to the output-amplifier tubes during the pulses. In addition it must contain pulse-forming networks or other pulse-generating circuits. Typical power-amplifier stages are discussed in the next two articles, after which several driver circuits are described.

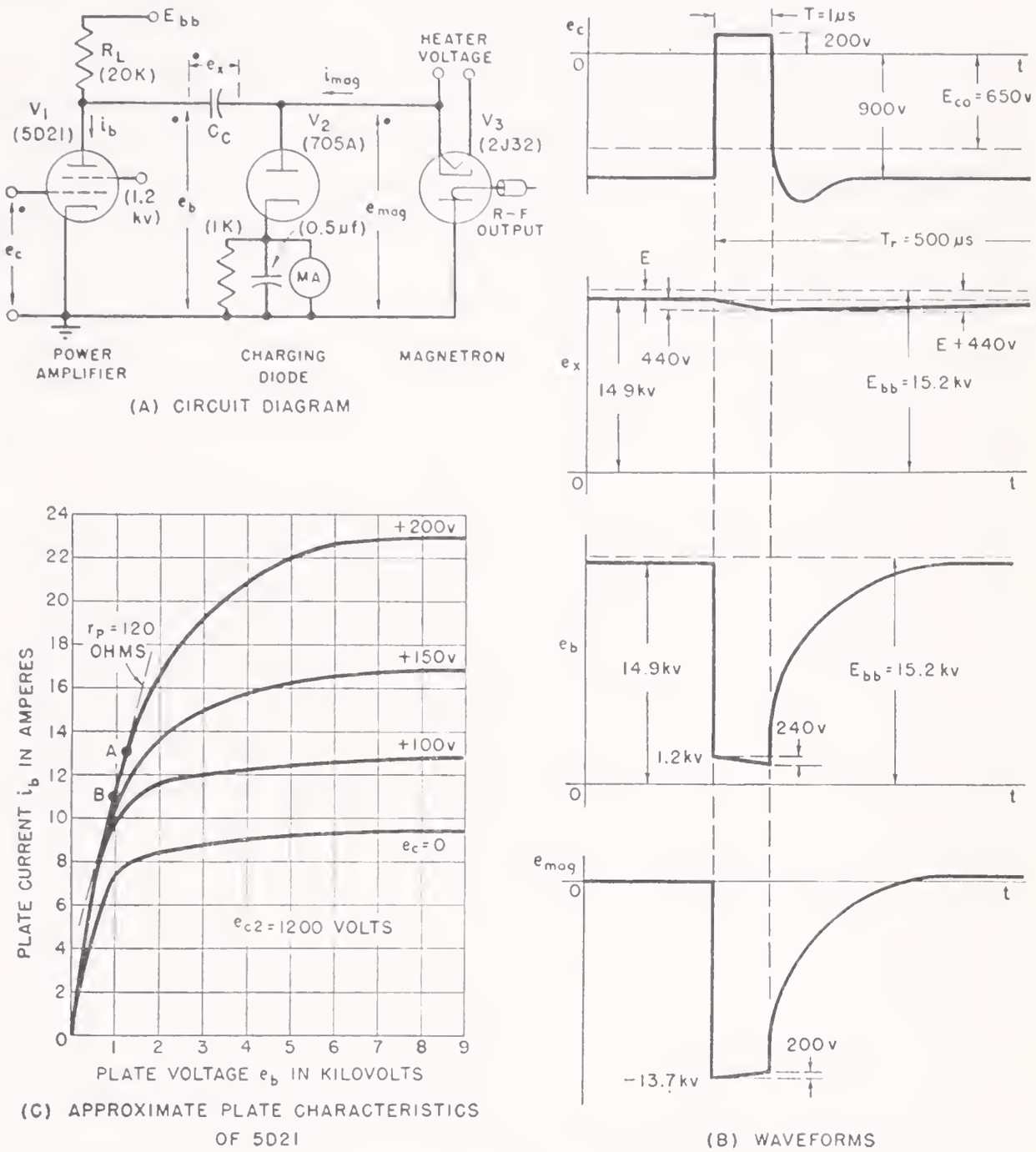


FIG. 25. *R-C*-coupled power amplifier for radar modulator.

10. *R-C*-coupled Power Amplifier. The basic form of *R-C*-coupled power amplifiers used in radar modulators is shown by the circuit of Fig. 25*A*, where the transmitting tube is assumed to be a magnetron. Because the grid voltage of the amplifier tube V_1 has the form of the e_c wave in Fig. 25*B*, V_1 is cutoff between pulses and conducts strongly during each pulse. Consequently, negative plate-voltage pulses are produced and

applied through capacitor C_c to the cathode of the magnetron (see waveforms of e_b and e_{mag} in Fig. 25*B*).

Operation of the *R-C* coupling circuit is in general the same as that of an *R-C* coupling circuit used between two overdriven amplifier tubes (see Art. 9, Chap. II). Two special details of the modulator circuit are important: (1) Because the magnetron can carry current in only one direction, a path for the charging current of C_c must be provided by an element in parallel with the magnetron. In Fig. 25*A*, diode V_2 allows C_c to charge between pulses; C_c discharges through the magnetron during the pulses. (2) The magnetron characteristics as well as the amplifier-tube characteristics are markedly nonlinear; therefore, plate characteristics of each tube are required to determine the operating conditions. Characteristics of the Type 2J32 magnetron are given on p. 795, and characteristics of the Type 5D21 amplifier tube appear in Fig. 25*C*.

Suppose that the magnetron is to be operated with a flux density of 1,500 gausses and a plate current of 12 amp at the beginning of a 1-μs pulse, and assume that the pulse repetition period is 500 μs. The magnetron characteristics show that an initial plate voltage of 13.7 kv is required and that the initial power output is approximately 75 kw.

During the pulse, capacitor C_c sends a discharge current through the magnetron and the amplifier-tube plate circuit, and simultaneously the power supply sends current through the amplifier tube (see the equivalent circuit in Fig. 26*A*). Since the voltage drop across R_L is very nearly equal to the voltage applied to the magnetron, the power-supply current is approximately 13.7/20 amp, or about 0.7 amp. Thus the current in the amplifier tube at the beginning of the pulse is 12.7 amp. Since e_c during the pulse is +200 volts, the initial operating point of the amplifier is point *A* in Fig. 25*C* and the initial value of e_b is 1.2 kv. The capacitor voltage e_x must therefore be 1.2 + 13.7 = 14.9 kv at the beginning of the pulse.

Discharge of capacitor C_c during the pulse causes the magnetron current and voltage to decrease. The capacitor must therefore be large enough to supply a current of nearly 12 amp for 1 μs without excessive change of voltage. Suppose that a decrease of magnetron current from 12 to 10 amp is allowable during the pulse. The corresponding allowable change of capacitor voltage may be determined by calculating the changes in amplifier-tube voltage and magnetron voltage. The amplifier-tube current is 10.7 amp at the end of the pulse (the change of the power-supply current being negligible), and the operating point at the end of the pulse is at *B* in Fig. 25*C*. The change of e_b between points *A* and *B* is somewhat difficult to determine by direct reading from the curves because of its small size. A more accurate procedure is to use the relation $\Delta e_b = r_p \Delta i_b$. The dynamic plate resistance r_p can be deter-

mined from the slope of the dotted line in Fig. 25*C* and is 120 ohms. Thus the 2-amp decrease of i_b is accompanied by a 240-volt decrease of e_b. The change of magnetron voltage is best determined in the same way. From the slope of the 1,500-gauss line, r_p for the magnetron is found to be 100 ohms, so that the change of e_{mag} is 200 volts. A decrease of capacitor voltage of 240 + 200 = 440 volts is therefore allowable. The average discharge current for the capacitor is (12 + 10)/2 = 11 amp, and the charge removed during the 1-μs pulse is 11 × 1 = 11 μcoulombs. The capacitance of C_c, therefore, should not be less than $^{11}/_{440}$ = 0.025 μf.

Between pulses, the equivalent circuit of Fig. 26*B* applies and the capacitor charges. If a 0.025-μf capacitor is used, the charging time constant is 20,000 × 0.025 = 500 μs because the plate resistance of the

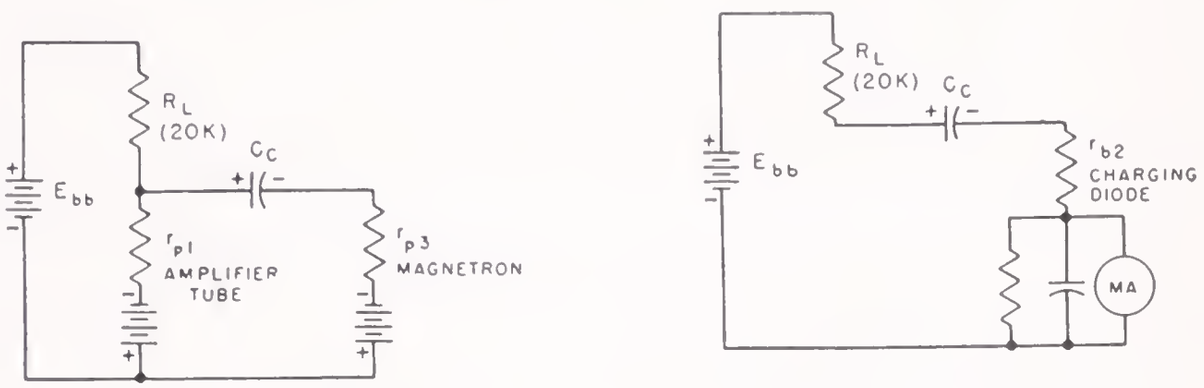


FIG. 26. Equivalent circuits for power-amplifier of Fig. 25*A* during and between pulses.

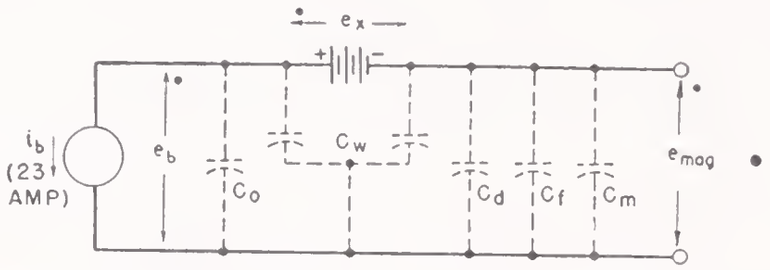
charging diode is negligible compared with 20,000 ohms. Therefore, as indicated on the waveform for e_x in Fig. 25*B*,

$$E = (440 + E)\epsilon^{-(500-1)/500}$$

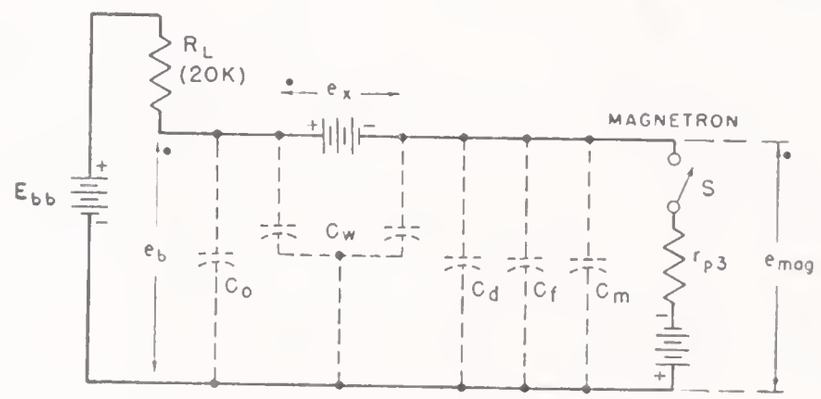
where E is the difference between E_{bb} and the capacitor voltage at the end of the charging period and 500 μs is the repetition period of the system. Thus E = 256 volts, and the plate voltage required for the specified operating conditions is 256 + 14,900 volts or very nearly 15.2 kv.

The steepness of the leading and trailing edges of the magnetron voltage pulse depends upon the rate at which the shunt capacitances of the circuit can be charged and discharged. These capacitances are indicated in Fig. 27 and include the following components: C_o, the output capacitance of the amplifier tube; C_w, the wiring capacitance and capacitance from C_c to ground; C_d, the diode capacitance; C_f, the capacitance between windings of the magnetron heater transformer; and C_m, the magnetron capacitance. The total shunt capacitance C_t is $C_o + C_w + C_d + C_f + C_m$ and for most modulators has a value between 50 and 120 $\mu\mu$f.

An approximate equivalent circuit applicable at the time of the leading edge of the pulse is that of Fig. 27*A*. During the short time of the pulse edge, there is almost no change of e_x; therefore, capacitance C_c is represented as a battery. The plate current of V_1, in the pulse-edge interval, is roughly constant at about 23 amp, as may be judged from the estimated position of the curve for $e_c = +200$ volts and e_b in the range of 1.2 to 14.9 kv in Fig. 25*C*. Thus tube V_1 is represented by a current source in the equivalent circuit. Current in R_L is neglected because it is smaller than the uncertainty in the value of the plate current. The static



(A) EQUIVALENT CIRCUIT AT BEGINNING OF PULSE



(B) EQUIVALENT CIRCUIT AT END OF PULSE

FIG. 27. Equivalent circuits for power amplifier of Fig. 25*A* at times of pulse edges.

characteristic of the magnetron (p. 795) indicates a negligible plate current during the major part of the pulse edge corresponding to magnetron voltages less than 12 kv. If the pulse edge is very steep, the dynamic magnetron characteristic does not exactly follow the static curve because r-f oscillations require time to start (Art. 8, Chap. XI), and little plate current flows before oscillations begin. Thus the magnetron is represented in Fig. 27*A* as an open circuit.

Within the accuracy limits of this equivalent circuit, e_b and e_{mag} decrease at a constant rate during the leading edge of the pulse because the constant current $i_b = 23$ amp flows through the capacitors. If the total capacitance is $C_t = 100$ $\mu\mu$f, the rate of decrease of voltage is

$$\frac{i_b}{C_t} = \frac{23}{100} \times 10^{-6} = 230{,}000 \text{ volts per } \mu\text{s}$$

and the time required for the total change of voltage $13.7/230 \approx 0.06$ μs. This interval is too short to be noticeable in the waveforms in Fig. 25*B*.

When the trailing edge of the pulse occurs, the applicable equivalent circuit is the one in Fig. 27*B*. Capacitances are the same as in Fig. 27*A*, but the plate current of tube V_1 is zero. At the beginning of the trailing edge, switch S in the magnetron equivalent circuit is closed, and e_b and e_{mag} rise very suddenly because of magnetron current. The sudden rise is short, because magnetron current becomes small, and switch S is effectively opened when the magnetron plate voltage reaches 12 kv. Further rise of the voltage follows an exponential curve as C_t receives current through R_L, and the time constant is $20{,}000 \times (100 \times 10^{-6}) = 2$ μs. (The rise of the voltages is actually slightly faster than this exponential because the magnetron current, though small, is not zero.) Because the trailing edge of the pulse is spread over a considerable time interval, the magnetron produces noise (see Art. 11, Chap. XI) for about 2 μs after the end of the pulse, and this noise interferes with the detection of objects at short range.

Use of Inductor to Increase Steepness of Trailing Edge of Pulse. To increase the steepness of the trailing edge of the magnetron voltage pulse, an inductor is often connected in parallel with the transmitter tube. This connection is shown in Fig. 28*A*, where L is the inductor. During the pulse, the negative voltage e_{mag} is applied across the inductor and causes current to build up in the direction of the arrow shown on the figure. The inductance current continues to flow for a short period after the pulse and serves as a discharge current for the capacitance shunted across the magnetron, thus causing the magnetron voltage to reach zero more quickly. The effect is illustrated by the trailing edges of the pulses in Fig. 28*B*, where the top waveform applies if the inductor is omitted, and the two other waveforms apply if the inductor is used.

When the inductor is used, the diode is no longer needed to complete the charging path for capacitor C_c, but it is still included in the circuit as a damping device. If it were omitted, oscillations would occur as shown in Fig. 28*B*, because inductance L and the shunt capacitance form a resonant circuit. Radio-frequency oscillations and noise would be produced by the magnetron during each negative swing of e_{mag}. The diode begins to conduct as soon as the voltage e_{mag} becomes positive, and the low resistance of the diode shunted across the L-C circuit changes the oscillations of e_{mag} into an overdamped R-L-C transient. As indicated in the figure, no negative overswing is obtained.

A disadvantage of using the inductor is that the change of magnetron current and voltage during the pulse is increased. The main reason for this effect is that the current built up in the inductor increases the ampli-

fier-tube current. As a result, the amplifier-tube voltage increases during the pulse, and less of the capacitor voltage is available for the magnetron.

If the current built up in the inductor during the pulse is not too great, its effect of decreasing the magnetron voltage can be compensated by increasing the size of C_c. Assume that this compensation is possible and

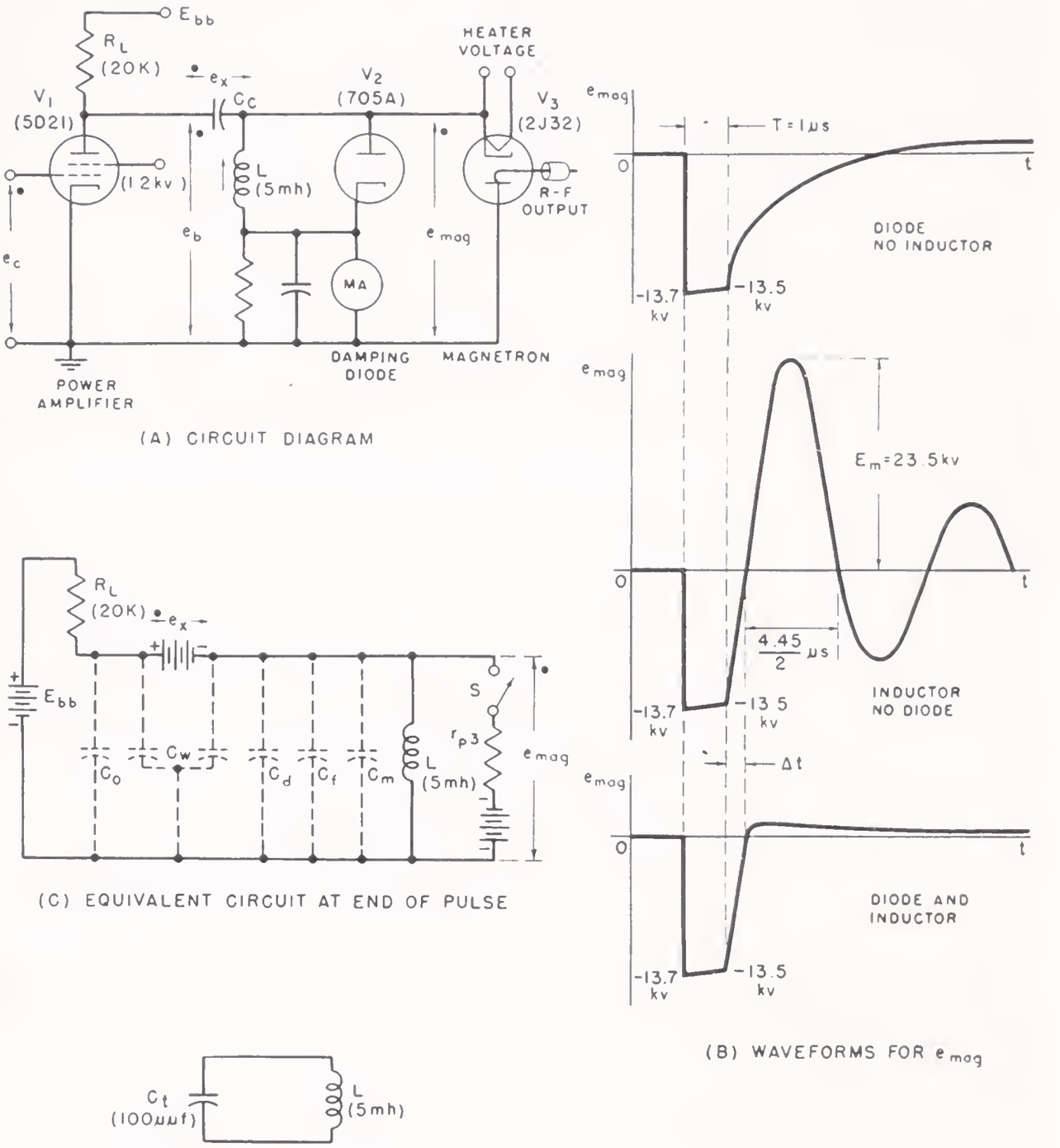


FIG. 28. Modulator power amplifier with inductor for increasing steepness of trailing edge of pulse.

that C_c and E_{bb} are adjusted so that during the entire pulse the magnetron operating conditions are the same for the circuit of Fig. 28*A* as for that of Fig. 25*A*. Initial values of magnetron voltage and current are then 13.7 kv and 12 amp. The initial current in the inductor is zero because the small value of this inductance (5 mh) allows the current in it to die

away completely between pulses. Consequently all initial currents and voltages are the same as in Fig. 25.

Operating conditions at the end of the pulse depend upon the amount of current built up in inductor L. This current can be determined from the inductance equation $e = L\,di/dt$, written in the form

$$\text{Average inductor voltage} = L \times \begin{pmatrix}\text{average rate of change}\\ \text{of inductor current}\end{pmatrix}$$

To duplicate the magnetron operating conditions of Fig. 25*A*, the average inductor voltage must be $13.7 - 0.2/2 = 13.6$ kv, and therefore the average rate of increase of inductor current is $13.6/5 = 2.72$ amp per μs. Thus the inductor current increases from zero to 2.72 amp during the 1-μs pulse.

The change of amplifier-tube current is the result of this *increase* of 2.72 amp together with the *decrease* of magnetron current of 2 amp. The amplifier-tube current therefore increases by $2.72 - 2 = 0.72$ amp, and the voltage e_b increases by $r_{p1}\Delta i_b = 120 \times 0.72 = 86.4$ volts. Since the magnetron voltage decreases by 200 volts during the pulse, the net decrease in capacitor voltage is $200 - 86.4 = 114$ volts. The average discharge current for the capacitor is

$$\tfrac{1}{2}[12 + (10 + 2.72)] = 12.36 \text{ amp}$$

because the magnetron current plus the inductor current flows through the capacitor. The charge removed during the 1-μs pulse is thus 12.36 μcoulombs. To obtain the assumed operating conditions, therefore, C_c must be $12.36/114 = 0.108$ μf—a value much larger than the 0.025-μf capacitance required when an inductor is not used.

Too small a value of L may make it impossible to limit the change of magnetron current to 2 amp, no matter how large C_c is made. For example, if L were 3 mh, e_b would be required to increase by 305 volts in order to maintain the assumed conditions of magnetron operation. Since the magnetron voltage decreases by only 200 volts, and since the capacitor voltage cannot increase, such a change of e_b is impossible. The magnetron current must therefore decrease to less than 10 amp during the pulse.

The effect of L in steepening the trailing edge of the pulse may be seen from an approximate calculation of Δt, the time interval between cutoff of V_1 and the return of e_{mag} to zero (see Fig. 28*B*). Since the diode V_2 does not conduct when e_{mag} is negative, Δt is the same for the two lower waveforms in Fig. 28*B*, and calculations may be based on the oscillation in the waveform obtained without the diode. The oscillation is caused by resonance between L and the shunt capacitances, as shown by the

equivalent circuit in Fig. 28*C*. The essentially constant voltage e_x across the battery representing C_c has no effect upon the oscillation; therefore the battery is omitted from the simplified circuit of Fig. 28*D*. As an approximation, currents in R_L and the magnetron are also ignored because they serve only to damp the oscillation slightly. If the total shunt capacitance is 100 $\mu\mu$f, the period of the oscillation is

$$2\pi \sqrt{100 \times 5 \times 10^{-3}} = 4.45 \ \mu s$$

The approximate amplitude of the oscillation may be determined from the energy relation in the circuit of Fig. 28*D*. At the instant that V_1 is cutoff, the current in L is 2.72 amp and the voltage across the shunt capacitance is 13.5 kv. The total energy stored in the resonant circuit at this instant is therefore

$$\tfrac{1}{2}Li^2 + \tfrac{1}{2}Ce^2 = \tfrac{1}{2}(5 \times 10^{-3}) \times (2.72)^2 + \tfrac{1}{2}(100 \times 10^{-12}) \times (13.5 \times 10^3)^2$$

$$= 0.0276 \text{ joule}$$

Except for the effects of damping, this energy is all transferred to the capacitor when the peak of the oscillation is reached. Thus

$$\tfrac{1}{2}(100 \times 10^{-12}) \times E_m^2 = 0.0276$$

where E_m is the peak voltage of the oscillation, and $E_m = 23.5$ kv.

From the peak voltage and period of the oscillation, the time Δt can be found. If θ represents the change of phase of the sine wave during the interval Δt, then sin θ is 13.5/23.5 = 0.574, θ is 35 deg, and Δt is

$$4.45 \times {}^{35}\!/_{360} = 0.43 \ \mu s$$

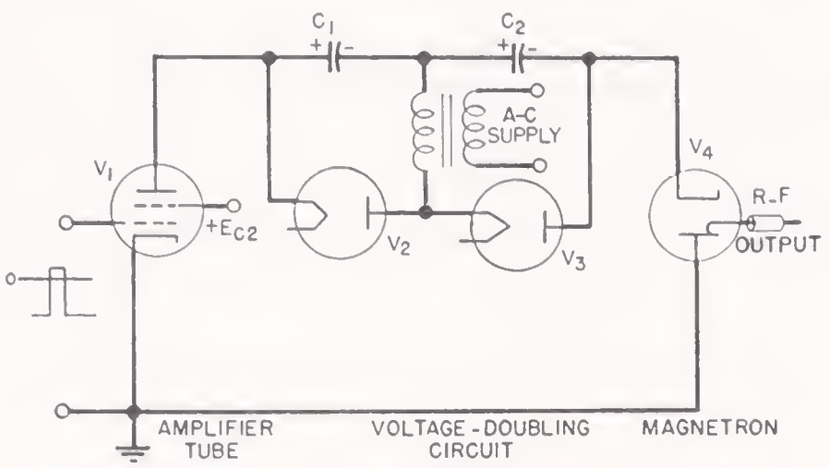
The inductor reduces the time interval in which noise is produced from about 2 μs to less than $\frac{1}{2}$ μs.

The very high value 23.5 kv found for E_m should be noted. If the damping diode were omitted, a positive peak voltage nearly this large would be produced. Since the voltage across C_c is nearly 15 kv, the plate voltage of V_1 would reach a peak value of nearly 23.5 + 15 = 38.5 kv, which is far above 20 kv, the allowable maximum voltage of the tube. The damping diode, therefore, is necessary not only to prevent unwanted radiation but also to prevent excessive voltages.

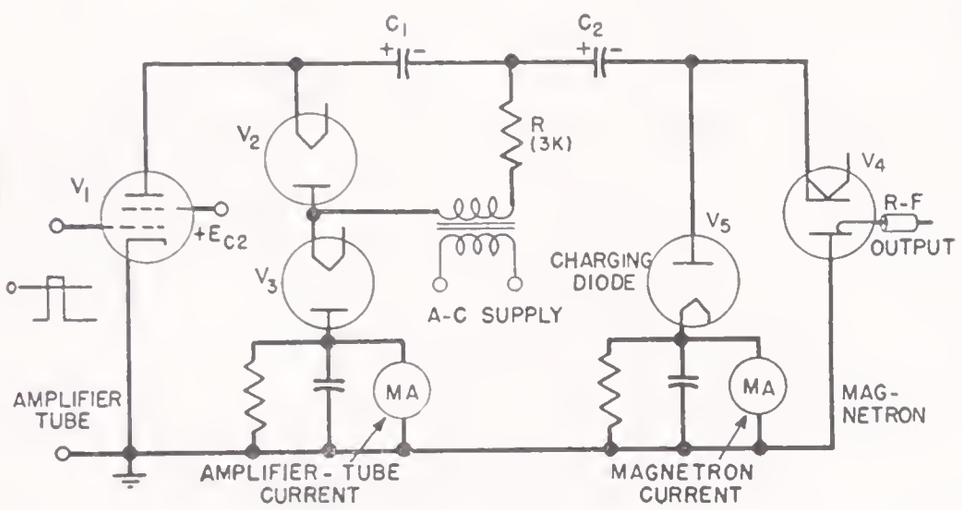
Variations of the Circuit. The plate-load impedance for the amplifier tube is often not the resistance R_L shown in Figs. 25*A* and 28*A*. Very frequently circuit diagrams show in this position a wire-wound 10-K resistor having an inductance of about 60 mh. Such a resistor limits the power-supply current during the pulse to much lower values than would a noninductive resistance of 10,000 ohms. In some circuits a small

inductance (5 to 10 mh) is used as plate load impedance to aid in making steep the trailing edge of the pulse. A resistance of 1,000 ohms or less is sometimes placed in series with the inductance. Such circuits allow rather large power-supply currents to flow during the pulse.

Instead of the diode in Fig. 25*A* or diode and inductor in Fig. 28*A*, a wire-wound resistance of 5 to 10 K with an inductance of 25 to 60 mh may



(A) DIRECT-COUPLED CIRCUIT



(B) MODIFICATION OF DIRECT-COUPLED CIRCUIT INCLUDING MILLIAMMETERS

FIG. 29. Circuits in which rectifier filter capacitors serve as coupling capacitor.

be connected across the magnetron. This impedance furnishes a charging path for C_c; at the same time the inductance provides some effect in steepening the pulse edge, and the resistance serves to damp the oscillations that follow the pulse. This circuit eliminates one tube at some sacrifice in performance.

The coupling capacitor C_c and the filter capacitor for the high-voltage power supply must have high-voltage ratings and rather large capacitance. To save space and weight, the power supply is sometimes

arranged so that the filter capacitors serve also as the coupling capacitor. A simple circuit for accomplishing this purpose is the direct-coupled one of Fig. 29*A*, in which the voltage-doubling rectifier, amplifier tube, and magnetron are connected in series. To permit measurement of the magnetron current without disconnecting either the magnetron plate or the amplifier-tube cathode from ground, the circuit may be modified, as in Fig. 29*B*, by connecting the plate of V_3 to ground through a by-passed milliammeter instead of to the magnetron cathode. This connection requires the addition of diode V_5 to provide a charging path for C_2 and of resistor R to limit the discharge current of C_1 through V_1 and V_3 during the pulse.

Power-amplifier Tubes. A list of the tubes commonly employed as power amplifiers in radar modulators is given in Table 2. The maximum ratings and typical operating conditions of the tubes are indicated. From data in the table, currents and voltages in modulators using these tubes can be estimated. For example, if a 13-kv, 14-amp pulse is to be applied to a magnetron by a modulator using an *R-C*-coupled Type 715B power amplifier, the plate voltage of the 715B would be about

$$1{,}250 \times 14/12.5 = 1{,}400 \text{ volts}$$

and a plate supply voltage of (13 + 1.4) kv, or about 14.5 kv, would be required.

TABLE 2. POWER-AMPLIFIER TUBES FOR MODULATORS

Type No.	*Kind of tube*	*Maximum ratings*		*Typical operating conditions*				
		Plate voltage, kv	*Peak plate current, amp*	e_{c2}, *volts*	*Between pulses,* e_{c1}, *volts*	*During pulse*		
						e_{c1}, *volts*	e_b, *volts*	i_b, *amp*
304TL	Triode	15	..	.....	−2,500	+50	800	1.6
304TH	Triode	15	..	.....	−1,000	+250	1,500	7.5
701A	Tetrode	15	15	1,000	−850	0	1,000	6.7
715A	Tetrode	14	10	1,200	−850	+200	1,000	10.0
715B	Tetrode	15	15	1,200	−850	+200	1,250	12.5
5D21	Tetrode	20	15	1,200	−850	+200	1,250	12.5
6C21	Triode	35	15	.....	−1,500	+1,500	2,500	15.5

11. Transformer-coupled Power Amplifier. When a considerable distance separates the transmitter from the modulator power-amplifier tubes, the pulse is conducted to the transmitter by a coaxial cable. Because the characteristic impedance of such a cable is usually 50 ohms, while the input impedance of the transmitter and optimum load for the

modulator are both of the order of 1,000 ohms, it is necessary to employ transformers to match these impedances.

A typical circuit for a modulator using a pulse cable and two pulse transformers is drawn in Fig. 30. In this circuit the transmitter tube is a Type 2J32 magnetron, and two Type 5D21 tetrodes are connected in parallel as the power amplifier. The small resistors in series with the grid and plate leads are necessary to prevent parasitic oscillations, that is, to prevent tubes V_1 and V_2 from functioning as a push-pull uhf oscillator (see Art. 5, Chap. X). The transformer connections are such that the negative pulse at the plate of V_1 and V_2 (produced by the positive grid-voltage pulse) causes a positive pulse at the center conductor of the pulse cable, and a negative pulse at the magnetron cathode.

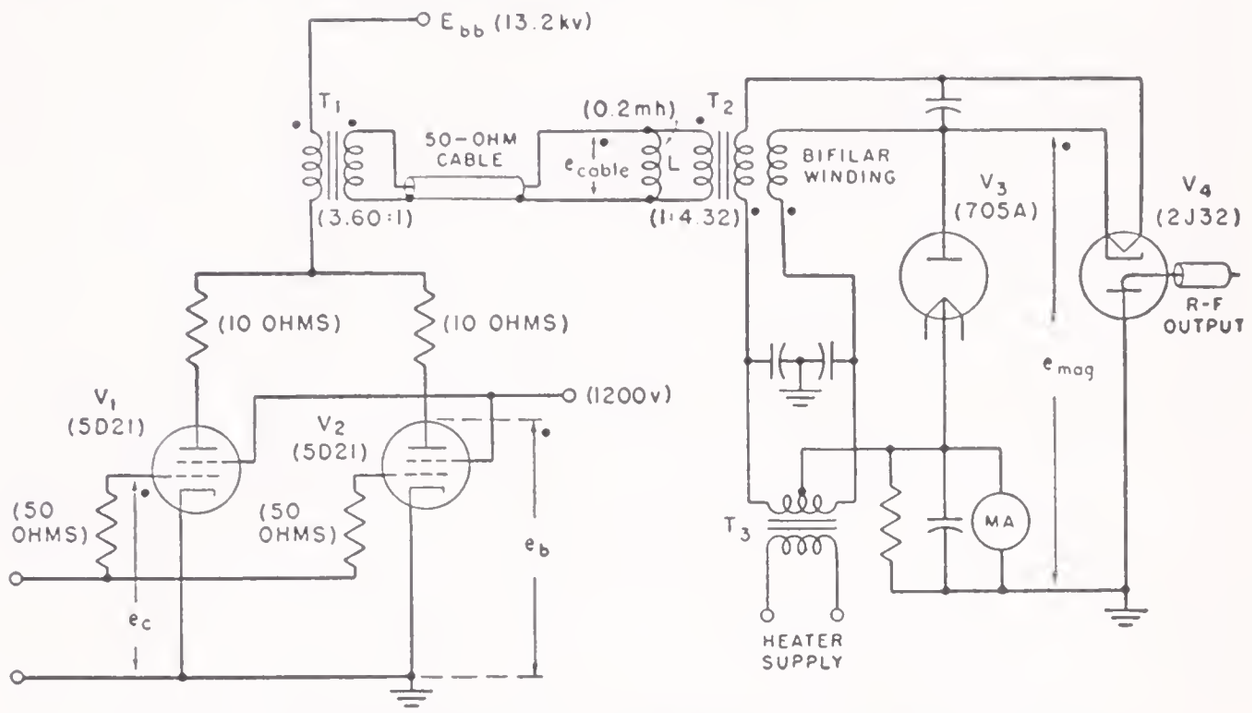


FIG. 30. Transformer-coupled power amplifier for radar modulator.

The circuit between pulse cable and magnetron in Fig. 30 is like the corresponding part of the line-pulsing modulator circuit in Fig. 18. Compare, for example, the bifilar transformer windings and the milliammeter circuits. The inductor and damping diode in each circuit prevent production of r-f output after the main pulse. In Fig. 30, the need for these elements arises because of high-frequency oscillations caused by leakage inductance of the transformer and shunt capacitance of the output circuit and because of reflections caused by mismatch of the pulse cable when the pulse begins and ends. Current built up in the inductor during the pulse causes the magnetron voltage to reverse at the end of the pulse, and the damping diode prevents oscillations from occurring in a circuit comprising the inductor and pulse-cable capacitance.

Suppose that the pulse shown as e_c in Fig. 31 is applied to the power-amplifier tubes and the modulator is intended to supply a magnetron

pulse having an initial voltage of 14 kv and an initial current of 15 amp. This operating point lies on the 1,500-gauss line of the magnetron characteristics (p. 795) and corresponds to a power output of 95 kw. The apparent load connected to the secondary of transformer T_2 is a resistance of 14,000/15 = 935 ohms. To terminate the cable properly, T_2 must therefore be a step-up transformer with a turns ratio of

$$\sqrt{935/50} = 4.32$$

The pulse voltage in the cable is 14/4.32 = 3.24 kv, and the pulse current is 15 × 4.32 = 64.8 amp.

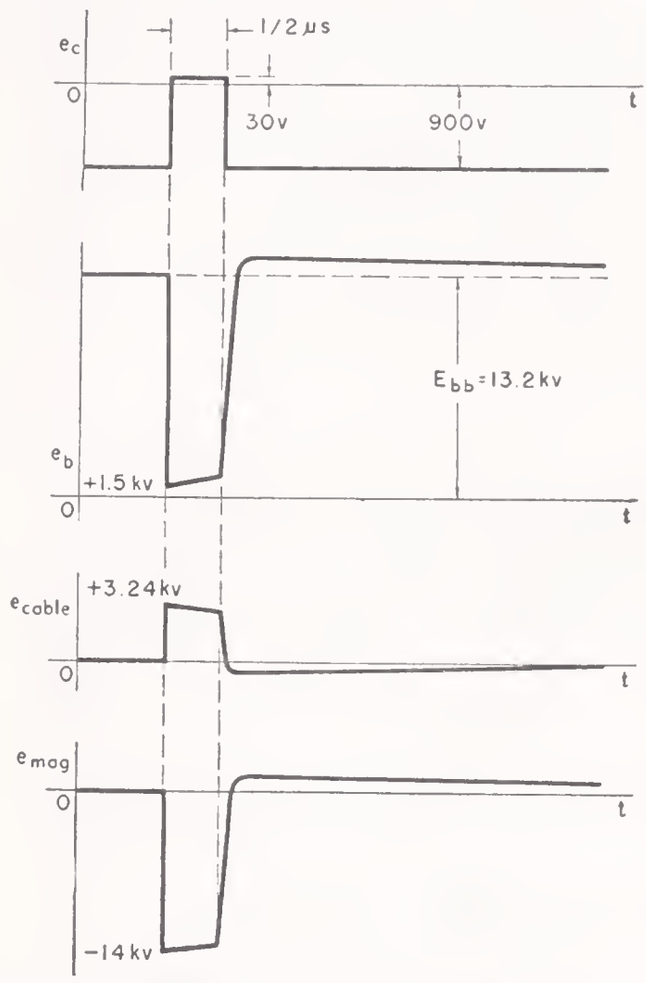


FIG. 31. Waveforms for power amplifier of Fig. 30.

The turns ratio of transformer T_1 determines the plate currents of tubes V_1 and V_2 during the pulse and fixes the value of E_{bb} needed to produce the specified magnetron pulse. If the plate current in each tube is to be 9 amp, a transformer-primary current of 2 × 9 = 18 amp is required and the step-down ratio of T_1 must be 64.8/18 = 3.60. The primary voltage of T_1 is then 3.60 × 3.24 = 11.65 kv. For i_b = 9 amp and e_c = +30 volts, the plate voltage e_b indicated by the 5D21 characteristics of Fig. 25C, p. 187, is 1.5 kv. Therefore E_{bb} must be

$$1.5 + 11.65 + 10 \times 9 \times 10^{-3} = 13.2 \text{ kv}$$

as indicated in Figs. 30 and 31. Note that reducing the step-down ratio of transformer T_1 would increase the amplifier-tube plate currents and decrease the required value of E_{bb}. Thus an excessively low ratio would require a tube current greater than the maximum allowable value of 15 amp, and an excessively high ratio would require a plate supply voltage of more than 20 kv, the maximum allowable value for the Type 5D21 tubes.

During the pulse, current builds up in the inductor L and in the magnetizing inductances of the transformers. This inductance current causes a slight mismatch at the termination of the pulse cable, and the exact determination of the change of operating conditions during the pulse is complicated by the reflection that results from this mismatch—

the two-way travel time in the cable is often of the same order of magnitude as the pulse duration. Because the magnetron current changes greatly with only a very small change of voltage, a very good approximation is obtained by assuming that during the pulse the increase of current in the inductances is exactly compensated by a decrease of magnetron current, so that the pulse cable is terminated in its characteristic resistance throughout the pulse. This is the same approximation made in Art. 8 for a similar circuit in a line-pulsing modulator.

The increase of current in the magnetizing inductance of T_1 causes the plate currents and voltages of amplifier tubes V_1 and V_2 to increase somewhat during the pulse, and corresponding decreases in e_{cable} and the magnitude of e_{mag} result as indicated in Fig. 31. Exact determination of these changes would be difficult because of reflections in the pulse cable. The magnetizing inductance of T_1 is, however, much larger than the inductance of L. Therefore, the magnetizing current is small, and its effect can be neglected in determining the change of magnetron operating conditions during the pulse.

Ordinarily the magnetizing inductance of T_2 also is large so that it may be neglected in parallel with the 0.2-mh inductance L. Because $e = L\,di/dt$ and e_{cable} has an essentially constant value of 3.24 kv during the pulse, the rate of change of current in L is (3.24 kv)/(0.2 mh) = 16.2 amp per μs, and at the end of the $\frac{1}{2}$-μs pulse the current in L is 16.2/2 = 8.1 amp. To compensate for this current, the magnetron current must decrease by 8.1/4.32 = 1.88 amp during the pulse. Note (from the 2J32 characteristics, p. 795) that the corresponding change of magnetron voltage is very small.

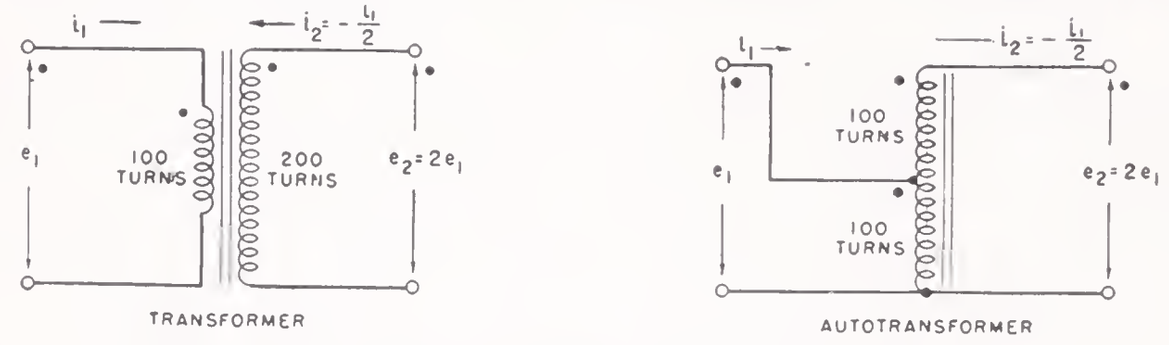
When tubes V_1 and V_2 are cut off at the end of the pulse, the current in L cannot stop immediately. Because of this current, the magnetron voltage decreases rapidly, reverses, and diode V_3 conducts. The form of the wave of e_{mag} after the pulse resembles closely that obtained in the modulator of Fig. 28, p. 192.

In these calculations, losses in the pulse cable and pulse transformer have been neglected. For this reason the power output indicated by the calculations (95 kw at the beginning of the pulse and 80 kw at the end of the pulse) is appreciably larger than may be expected from an actual circuit using the same tube, plate supply voltage, and input wave.

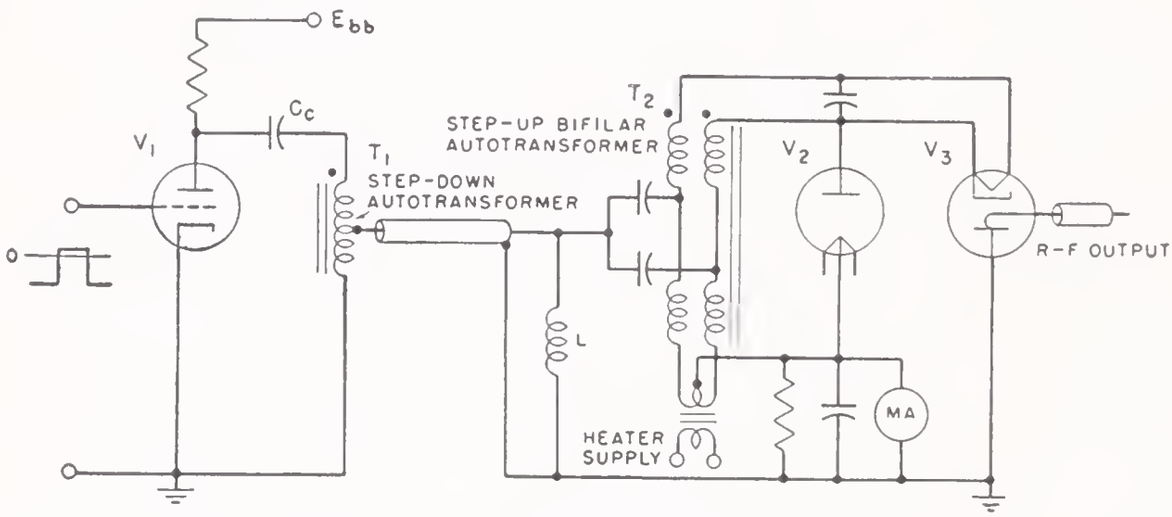
Space and weight may sometimes be saved by using autotransformers instead of two-winding transformers. In Fig. 32*A*, a transformer and an autotransformer having equivalent impedance-changing properties are shown, and in Fig. 32*B*, an application of autotransformers in a radar modulator is indicated. A coupling capacitor is required between power-amplifier tube and autotransformer to remove the direct component of the power-amplifier plate voltage. This arrangement is sometimes called

shunt feed. Because an autotransformer cannot invert the polarity of signals applied to it, the pulses at the center conductor of the cable are negative. The capacitors at the output of the cable prevent the autotransformer windings and inductance from forming a d-c short circuit across the milliammeter.

12. Bootstrap Driver. The driver stage of a driver-power-amplifier modulator must supply to the grid of the amplifier tube a large negative



(A) AUTOTRANSFORMER EQUIVALENT TO TWO-WINDING TRANSFORMER



(B) MODULATOR CIRCUIT EMPLOYING AUTOTRANSFORMER

FIG. 32. Use of autotransformers in radar modulator.

voltage between pulses (to maintain the tube cutoff) and a fairly large positive voltage during the pulse (to enable the tube to conduct large currents with low plate voltage). Values of grid-to-cathode voltage of -800 to $-2{,}500$ volts are required between pulses, and voltages as high as $+1{,}500$ volts may be needed during the pulse. Since the grid current during the pulse is of the order of an ampere, the driver stage is required to furnish a peak power of several hundred watts. The driver must control this power in such a way that the pulses produced have steep sides, flat tops, and a specified duration.

The earliest widely used circuit for producing nearly rectangular pulses of 1-μs duration is the *bootstrap driver.* In it the pulse is produced by

discharge of an artificial line through an argon thyratron. Because of the voltage limitations of the tube, the pulse amplitude obtainable is only a few hundred volts, and a driver amplifier is therefore required between the gas tube and the modulator power amplifier. The driver-amplifier tube, like the power-amplifier tube, must be cutoff between pulses and conducting during the pulse. (The reverse conditions of operation would require an unreasonably high-power tube and would make the driver-amplifier stage consume a power in excess of a kilowatt—more than is used by the final power amplifier.) If a conventional plate-loaded amplifier were employed, negative pulses would be obtained and an inverting transformer would be needed to supply positive pulses to the power-amplifier grid. When the bootstrap driver was devised, satisfactory transformers for use with 1-μs pulses were not available, and therefore the cathode-loaded amplifier, shown in Fig. 33, was used to produce positive pulses directly.

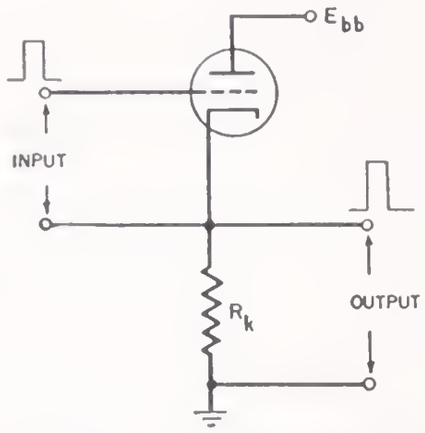


FIG. 33. Cathode-loaded amplifier or bootstrap circuit.

Note that this circuit is not a cathode follower because the input pulse is applied between grid and cathode; the voltage amplification equals that of a plate-loaded amplifier. The name *bootstrap* is applied to the cathode-loaded amplifier because an increase of grid voltage increases the cathode voltage and the cathode-voltage rise further increases the grid-to-ground voltage.

A complete bootstrap-driver circuit is drawn in Fig. 34, and the waveforms produced in it are shown in Fig. 35*B*. The grid-bias voltages for the gas tube V_2 and the driver-amplifier tube V_3 are adjusted so that both tubes are cutoff between pulses. The output voltage e_o between pulses is therefore $-1{,}100$ volts, a value capable of maintaining the power-amplifier tube cutoff. The period between pulses is sufficiently long to allow the pulse-forming network to charge to its final value of 600 volts and to allow capacitance C_1 to charge to 1,200 volts so that the plates and cathode of V_1 are at the same voltage.

The positive input pulse is applied through capacitor C_1 and diode V_1 to the grid of the gas tube V_2 and causes V_2 to conduct. Discharge of the artificial line through this tube in series with the parallel combination of R_{k2} and the grid resistance of tube V_3 causes a positive pulse to be applied to the grid of V_3, and plate current in this tube produces the output pulse (see waveforms of e_{c3} and e_o). Because of the bootstrap connection, the voltage of the entire circuit of tubes V_2 and V_3 rises 1,400 volts during this pulse. The bootstrap capacitor C_2 causes the grid voltage of V_2 to rise with the cathode voltage, thus preventing the flow of a large

positive-ion grid current that would damage the grid. Diode V_1 allows the rise of grid voltage to occur without effect upon the input circuit.

The equivalent circuit for the discharge of the pulse-forming network during the pulse is drawn in Fig. 35*A*. In it the pulse-forming network is represented by its characteristic resistance in series with a battery (the initial network voltage), the plate-to-cathode circuit of V_2 is represented as an 8-volt battery (because the voltage drop across the gas tube is nearly constant at this value), and the grid resistance of the two sections of tube V_3 in parallel (including the effect of the parasitic-suppressor

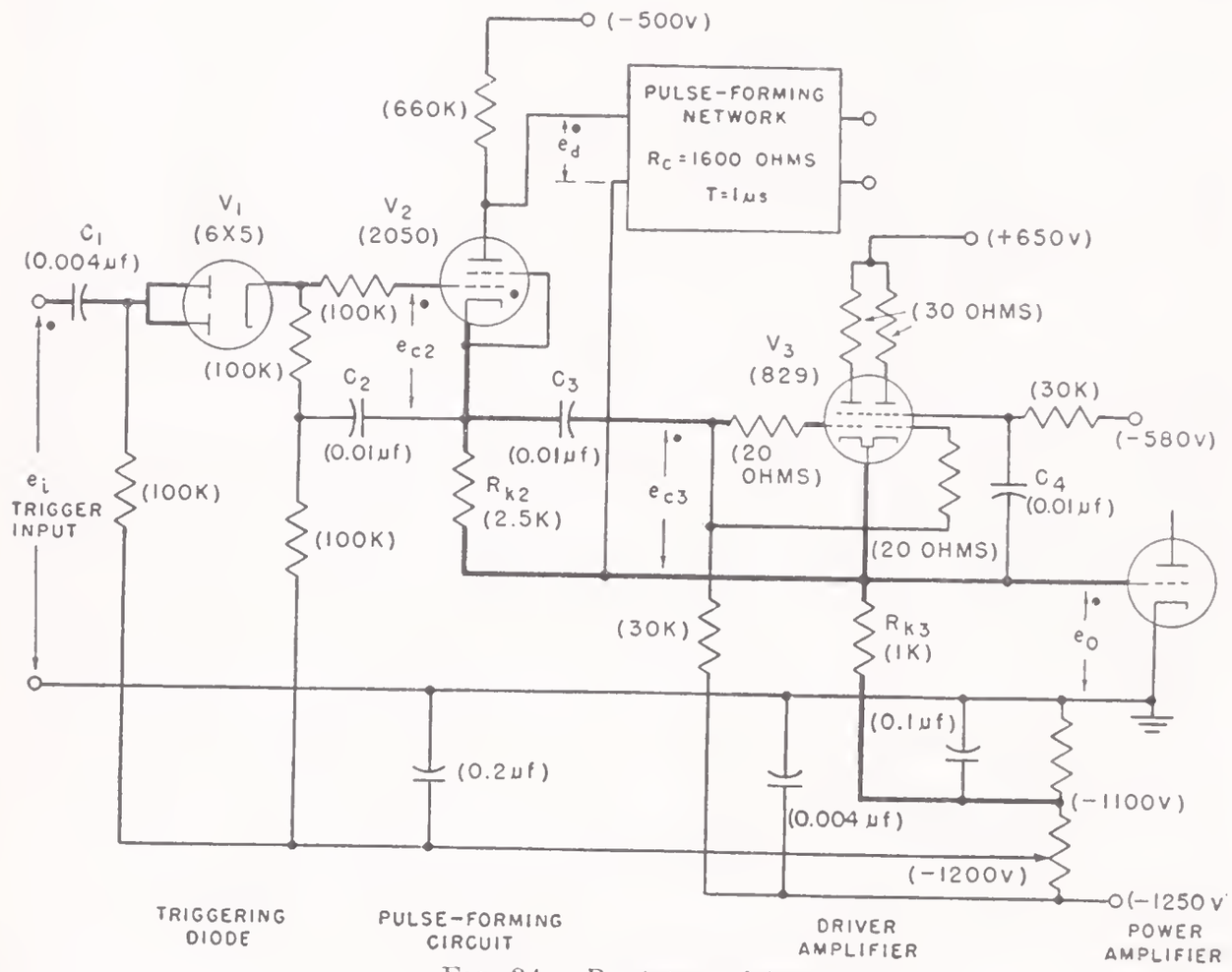


FIG. 34. Bootstrap driver.

resistors) is represented by a 100-ohm resistor r_{c3}. Currents in the 660-K plate-load resistor for V_2 and the 30-K grid-return resistor for V_2 are small and are neglected in the equivalent circuit. Application of Thévenin's theorem to the part of the circuit within the dotted rectangle allows calculation of voltages during the pulse; the values $e_d = 170$ volts and $e_{c3} = +20$ volts are obtained.

The voltage e_d drops from 600 to 170 volts at the beginning of the pulse and remains at 170 volts during a 1-μs interval corresponding to the two-way travel time of the 430-volt wavefront. At the end of the interval, e_d drops an additional 430 volts (see waveform of e_d), the effect being the same as if the network were discharged into a little less than one-half of

its characteristic resistance. The negative value reached by e_d stops conduction in V_2, and the gradual stepped-exponential rise of e_d from -260 to $+600$ volts begins immediately. The cathode load for V_2 must be chosen so that e_d becomes negative after the pulses in order that V_2 may have time to deionize after each pulse.

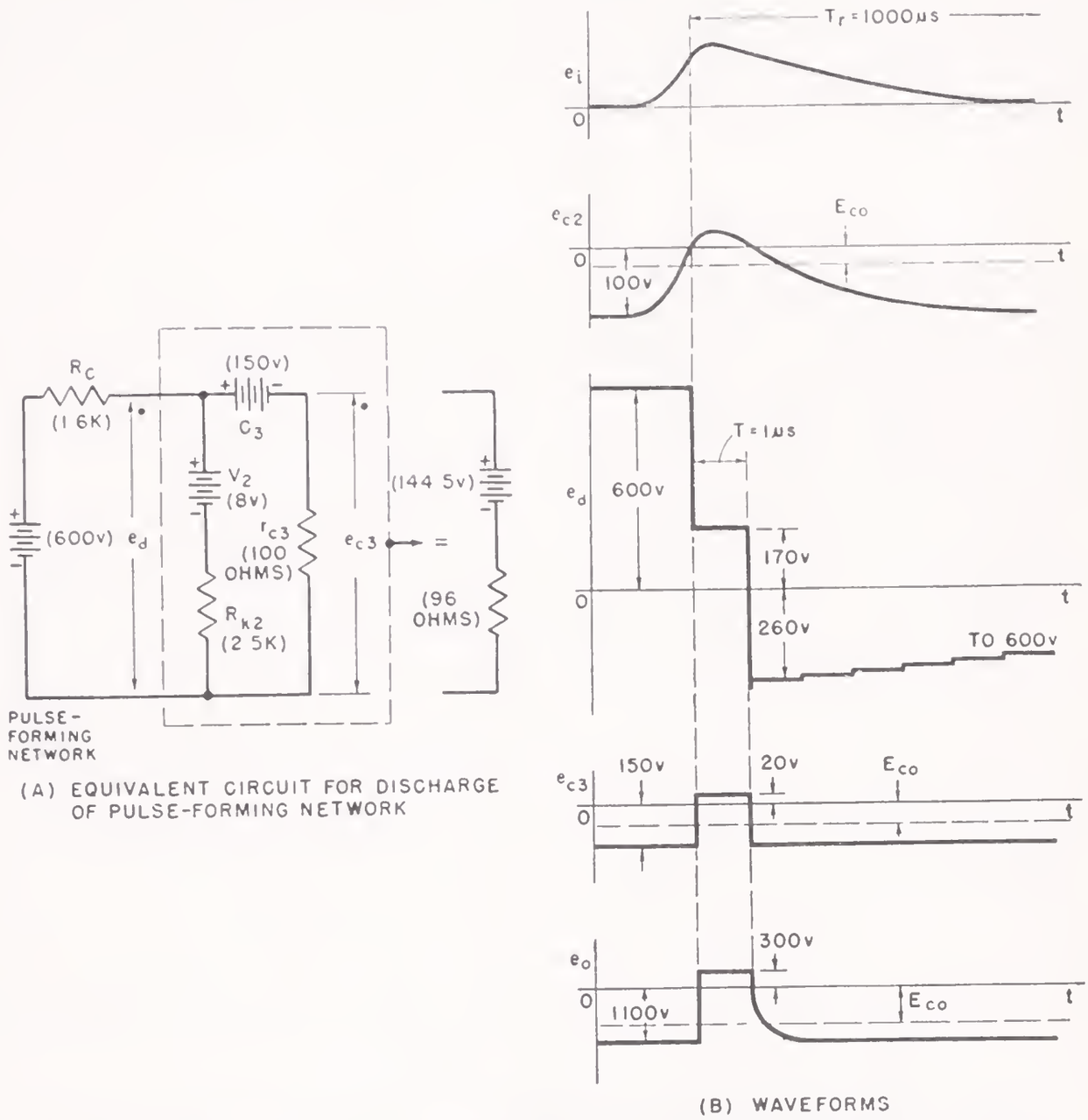


FIG. 35. Equivalent circuit and waveforms for the bootstrap driver of Fig. 34.

When the grid-to-cathode voltage of V_3 is raised to $+20$ volts, plate current from V_3 flows through R_{k3} and the grid resistance of the power-amplifier tube and raises the cathode voltage of V_3 from $-1,100$ to $+300$ volts. This rise of e_o not only raises the voltage of all points in the circuit of V_2 but also, because of capacitor C_4, raises the screen-grid voltage of V_3. Capacitor C_4 charges to about 400 volts between pulses and maintains this screen-grid-to-cathode voltage for V_3 during the pulse; discharge of the capacitor furnishes the screen-grid current.

To verify the +300-volt value for e_o during the pulse, the plate characteristics of the Type 829 tube and the grid characteristics of the power-amplifier tube are required. The bootstrap circuit ordinarily drives two Type 304TH triode power-amplifier tubes in parallel which, under the normal operating conditions, require a total grid current of 1.6 amp at 300 volts. (Actual grid characteristics rather than a value of static grid resistance are used because the grid current of these tubes is greatly affected by plate voltage.) The total plate current in the two sections of V_3 is the sum of this grid current and the current in R_{k3}, or

$$1.6 + \frac{1,400}{1,000} = 3.0 \text{ amp}$$

Plate characteristics of the Type 829 tube that include the effect of the parasitic-suppressor resistors in grid and plate loads (see Fig. 37*A*) show that e_b is 350 volts when e_c is +20 volts and the current in each section is 3.0/2 = 1.5 amp. The sum of the voltage drops in the plate circuit of V_3 is therefore 300 + 350 = 650 volts, which value agrees with the supply voltage indicated in Fig. 34.

At the end of the pulse, the cutoff of current in the gas tube V_2 returns the voltage e_{c3} to −150 volts, thus cutting off V_3 and causing e_o to decrease to −1,100 volts. As indicated in Fig. 35*B*, the decrease of e_o is not instantaneous but occurs at a rate limited by the shunt capacitance of the circuit. To speed the decrease, an inductance of about 2.5 mh is usually connected across the 1-K resistor R_{k3}. This inductor functions in the same manner as the shunt inductor used in the power amplifier and has the same undesirable effect of causing the top of the pulse to slope a little.

Sometimes a resistor and capacitor in parallel are placed between the cathode of V_2 and the connection to C_3. A voltage builds up across the capacitor during the pulse and dies away slowly after the pulse. The added positive voltage thus obtained at the cathode of V_2 helps to provide time for deionization and for the triggering voltage to die away before the plate-to-cathode voltage of V_2 becomes positive.

13. Line-controlled Blocking-oscillator Driver. Radar pulses as short as 0.1 μs or as long as several microseconds can be produced by means of the blocking-oscillator driver of Fig. 36. This driver has a somewhat simpler circuit than the bootstrap driver because the pulse transformer makes cathode coupling unnecessary and the blocking-oscillator connection eliminates the need for a gas tube. The blocking oscillator in Fig. 36 functions in the same manner as does the one described in Art. 21 of Chap. II, except that the pulse duration is controlled by a pulse-forming network instead of by a capacitor and the transformer magnetizing inductance. To trigger the blocking oscillator, pulses are applied in the grid circuit by a cathode follower, tube V_1. Positive out-

put pulses are applied to the power amplifier, tube V_3, by means of the third winding of the pulse transformer.

Waveforms for the blocking-oscillator driver are drawn in Fig. 37*B*. Between pulses the cathode voltage of V_1 is 15 volts, the voltage e_d across the pulse-forming network reaches a final value of 15 + 125 = 140 volts, and tubes V_2 and V_3 are cutoff because of the negative voltages at their grids. When positive trigger pulses are applied to the grid of V_1, the resulting rise of the cathode voltage e_k is coupled through the network and the transformer winding to V_2. The rise of e_{c2} above cutoff starts a sudden switching action in the blocking oscillator that ends with V_2

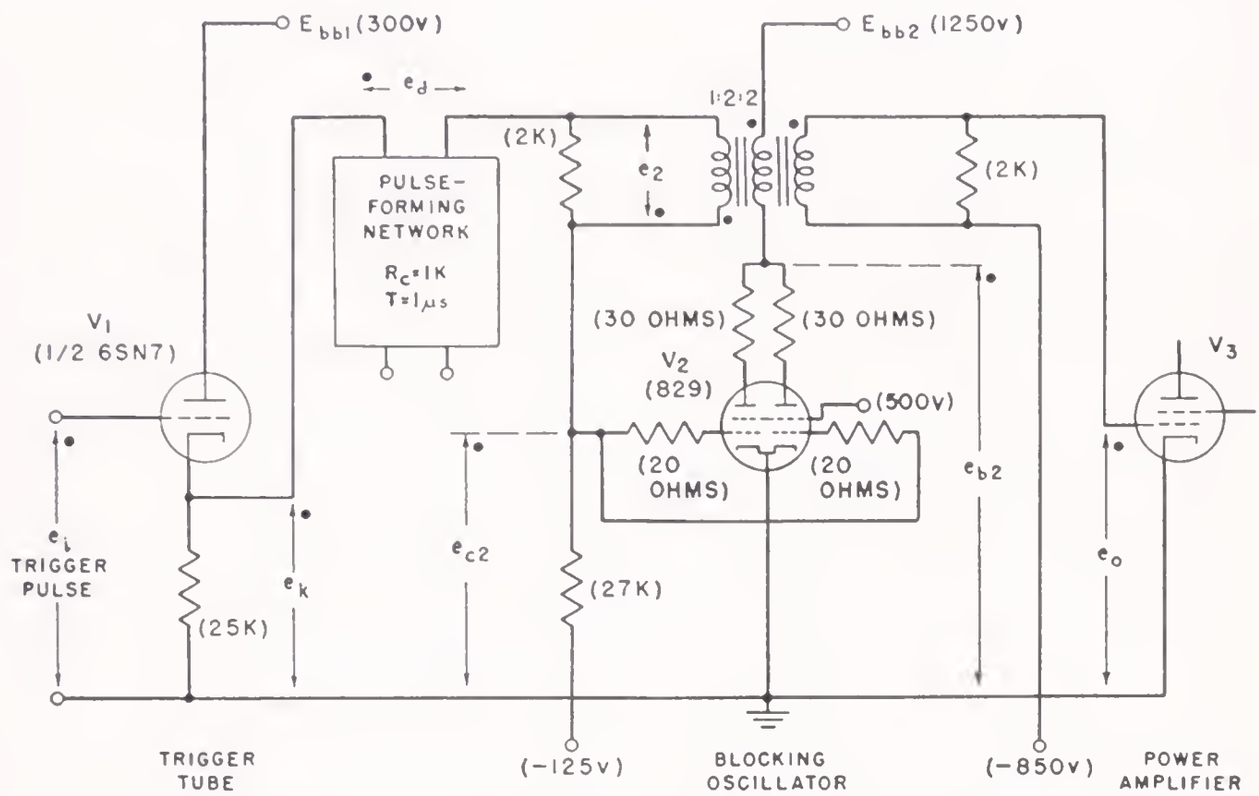
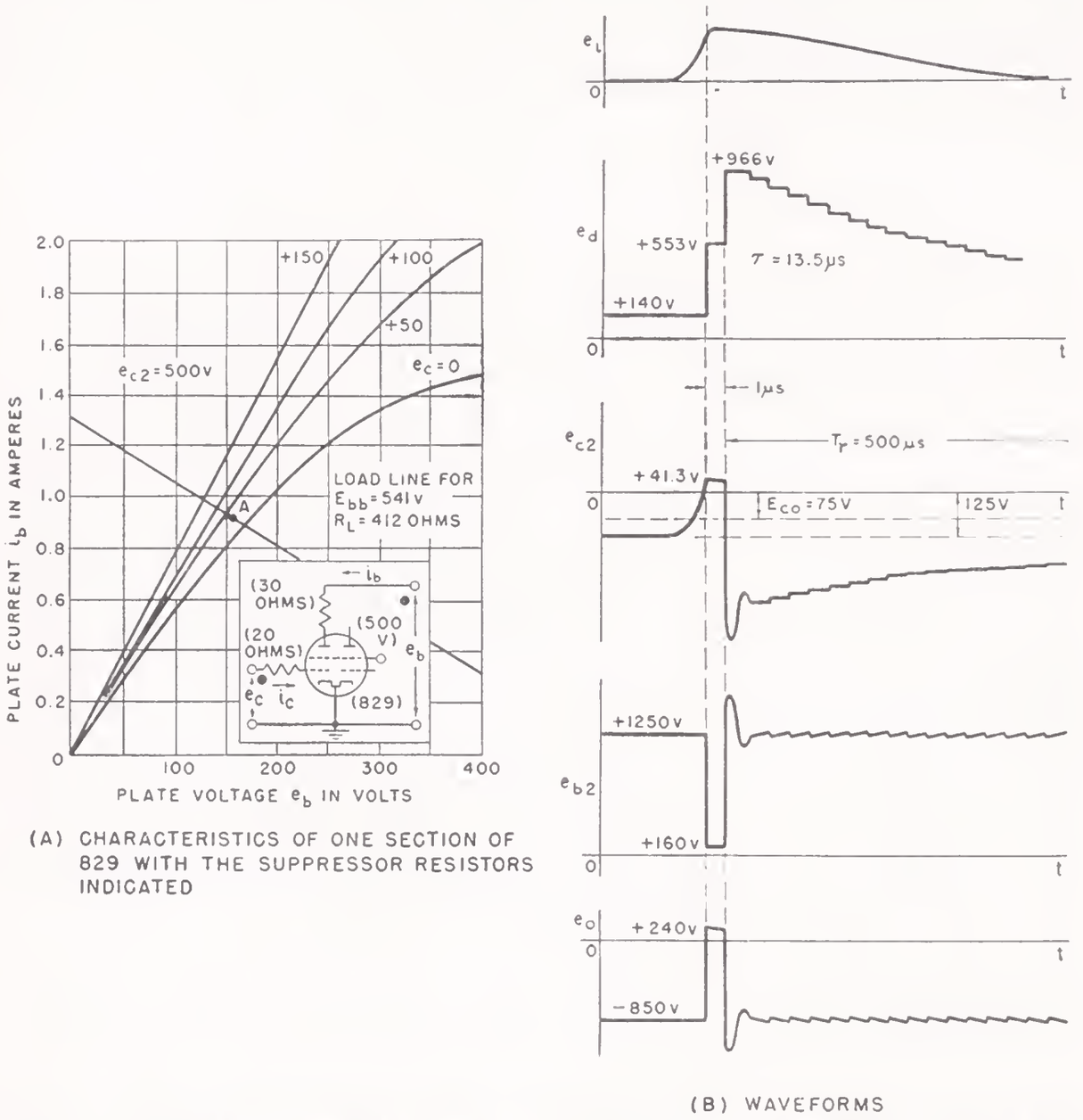


FIG. 36. Line-controlled blocking-oscillator driver.

conducting strongly. A negative pulse therefore appears at the plate of V_2, and positive pulses appear at the grid of V_2 and V_3. Grid current for V_2 flows through the pulse-forming network and causes a sudden increase of e_d. Consequently, if the network is an artificial line, a wavefront travels toward the open end, is reflected, and causes a second increase of e_d when it returns to the input end. This second change reduces the grid voltage of V_2 and starts a second switching action that ends the pulse. After the pulse, e_d returns to its original value in small steps, the 27-K grid-return resistor serving as a discharge path.

As a first step in the determination of the amplitudes of the pulses produced by the blocking oscillator, the equivalent circuits of Fig. 38 are drawn. Figure 38*A* shows the circuit elements coupled by the pulse transformer to the plates of V_2. The power-amplifier grid circuit is represented by a 250-ohm resistance r_{c3}, which is the static grid resistance

of the tube most frequently driven by the circuit (a Type 715B tetrode) for the usual operating conditions (1,200 to 1,400 volts at the screen grid and +200 to +300 volts at the control grid). Similarly, the grid circuit of both halves of V_2, including the suppressor resistors, is represented by $r_{c2} = 100$ ohms. The pulse-forming network appears as its characteristic resistance in series with a battery representing the initial network voltage.



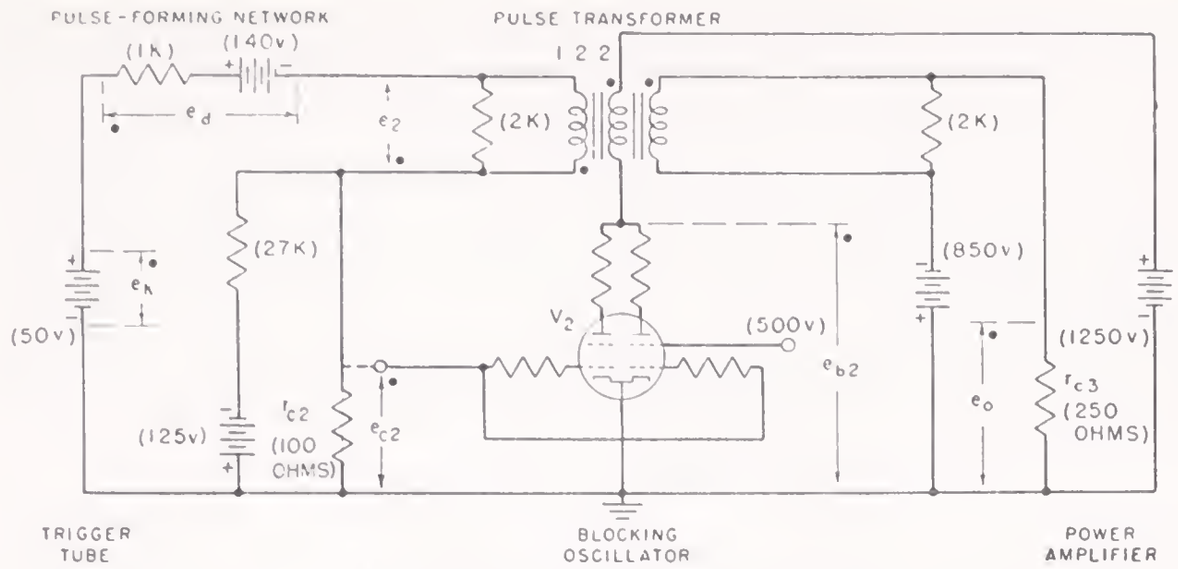
(A) CHARACTERISTICS OF ONE SECTION OF 829 WITH THE SUPPRESSOR RESISTORS INDICATED

(B) WAVEFORMS

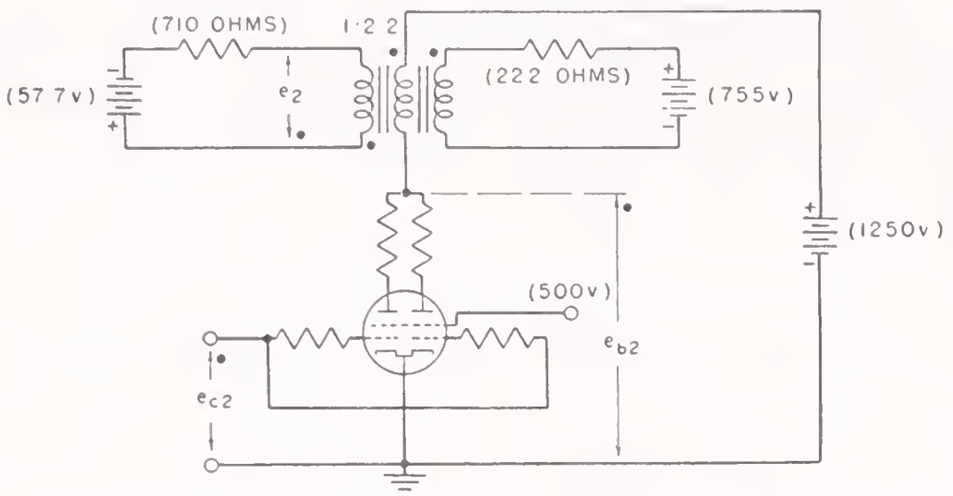
FIG. 37. Tube characteristics and waveforms for blocking-oscillator driver of Fig. 36.

The cathode-follower equivalent circuit depends upon the output impedance of the source of grid-cathode voltage and the rate of rise of the trigger-pulse voltage. Changes in the cathode follower have little effect on the output pulse of the blocking oscillator; therefore, for simplicity, the approximation is made that e_k has a constant value of +50 volts throughout the pulse. This value makes allowance for the reduction of e_k caused by current in the blocking-oscillator grid circuit.

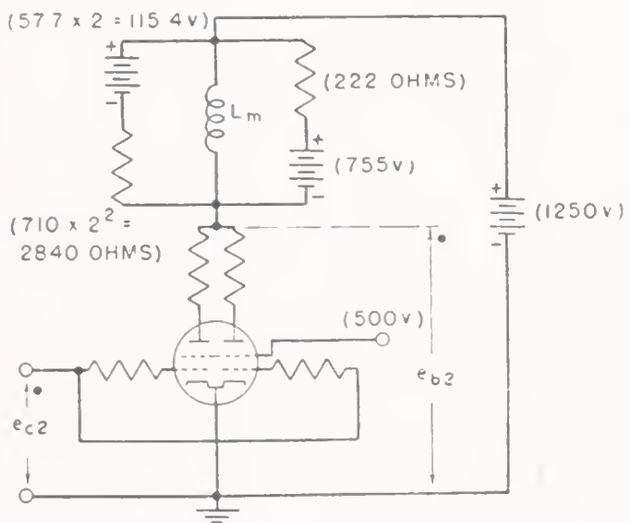
The circuits connected to the blocking-oscillator-grid winding and to the output winding of the transformer may each be simplified by use of



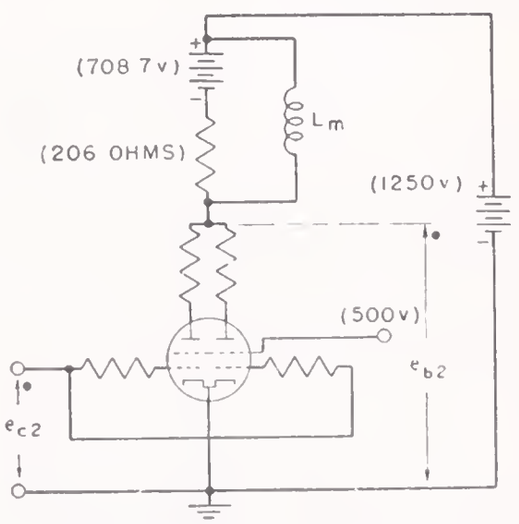
(A) EQUIVALENT CIRCUIT DURING PULSE



(B) SIMPLIFICATION OF FIG 38A OBTAINED BY USE OF THÉVENIN'S THEOREM



(C) SIMPLIFICATION OF FIG 38B OBTAINED BY REFERRING GRID CIRCUITS TO TRANSFORMER PRIMARY



(D) SIMPLIFICATION OF FIG 38C OBTAINED BY USE OF THÉVENIN'S THEOREM

FIG. 38. Equivalent circuits for blocking oscillator of Fig. 36.

Thévenin's theorem, as indicated in Fig. 38*B*. The result may be further simplified by referring these circuits to the plate winding of the transformer, as in Fig. 38*C*, the turns ratio being used to refer voltages and the square of the turns ratio to refer resistances. The transformer magnetizing inductance is marked L_m in Fig. 38*C*. The final simplification of the circuit, illustrated in Fig. 38*D*, is obtained by a second application of Thévenin's theorem.

When tube V_2 first begins to conduct, the current in L_m is zero, and therefore the circuit of Fig. 38*D* is equivalent to a battery of

$$1{,}250 - 708.7 = 541.3 \text{ volts}$$

applied to the tube through a resistance of 206 ohms. The effective load resistance for *each section* of tube V_2 is therefore $2 \times 206 = 412$ ohms, and the operating point lies on the load line drawn in Fig. 37*A*. The position of the operating point on this line depends upon the grid voltage e_{c2} and may be determined by a trial-and-error calculation. The final trial is $e_{c2} = 41.5$ volts, so that the operating point is *A* in Fig. 37*A*. The calculations are as follows: At point *A*, $e_{b2} = 160$ volts,

$$e_2 = \frac{1{,}250 - 160}{2} = 545 \text{ volts}$$

and from Fig. 38*A* (if the small current in the 27-K grid-return resistor is neglected)

$$e_{c2} = 100 \times \frac{545 - 140 + 50}{100 + 1{,}000} = +41.3 \text{ volts}$$

which checks the trial value of e_{c2}. The initial values of e_o and e_d may now be determined from the circuit of Fig. 38*A*:

$$e_o = (1{,}250 - 160) - 850 = +240 \text{ volts}$$

$$e_d = 140 + \frac{1{,}000}{100} e_{c2} = 553 \text{ volts}$$

During the pulse, conditions are constant except that (1) the plate current of V_2 increases slowly because of the growth of magnetizing current in the transformer and (2) an increase of e_i may be expected to occur, increasing the current delivered by the grid winding of the transformer and requiring an additional increase of plate current. As explained in Art. 21 of Chap. II, the blocking-oscillator tube supplies this increased current with little change of plate or grid voltage.

If the pulse-forming network is thought of as an artificial line, the rise of e_d at the beginning of the pulse ($553 - 140 = 413$ volts) is the voltage of a wavefront that travels to the open end of the line and back during a 1-μs interval. Therefore, after 1 μs, e_d rises by a second 413-volt step, and

the line is statically charged to $140 + 413 + 413 = 966$ volts. As a result of the second rise of e_d, the grid voltage e_{c2} drops suddenly, and the pulse is ended by a switching action. The pulse-forming network begins immediately to discharge in small steps toward its original voltage of 140 volts. Since the storage capacitance of the line (see Eq. 1, p. 164) is $[10^{-6}/(2 \times 10^3)] \times 10^{12} = 500$ $\mu\mu$f, the time constant of the exponential approximated by the discharge curve is $0.027 \times 500 = 13.5$ μs; and the discharge is essentially complete in a fraction of the repetition period.

The current built up in the transformer magnetizing inductance during the pulse must flow into the shunt capacitances of the circuit when the

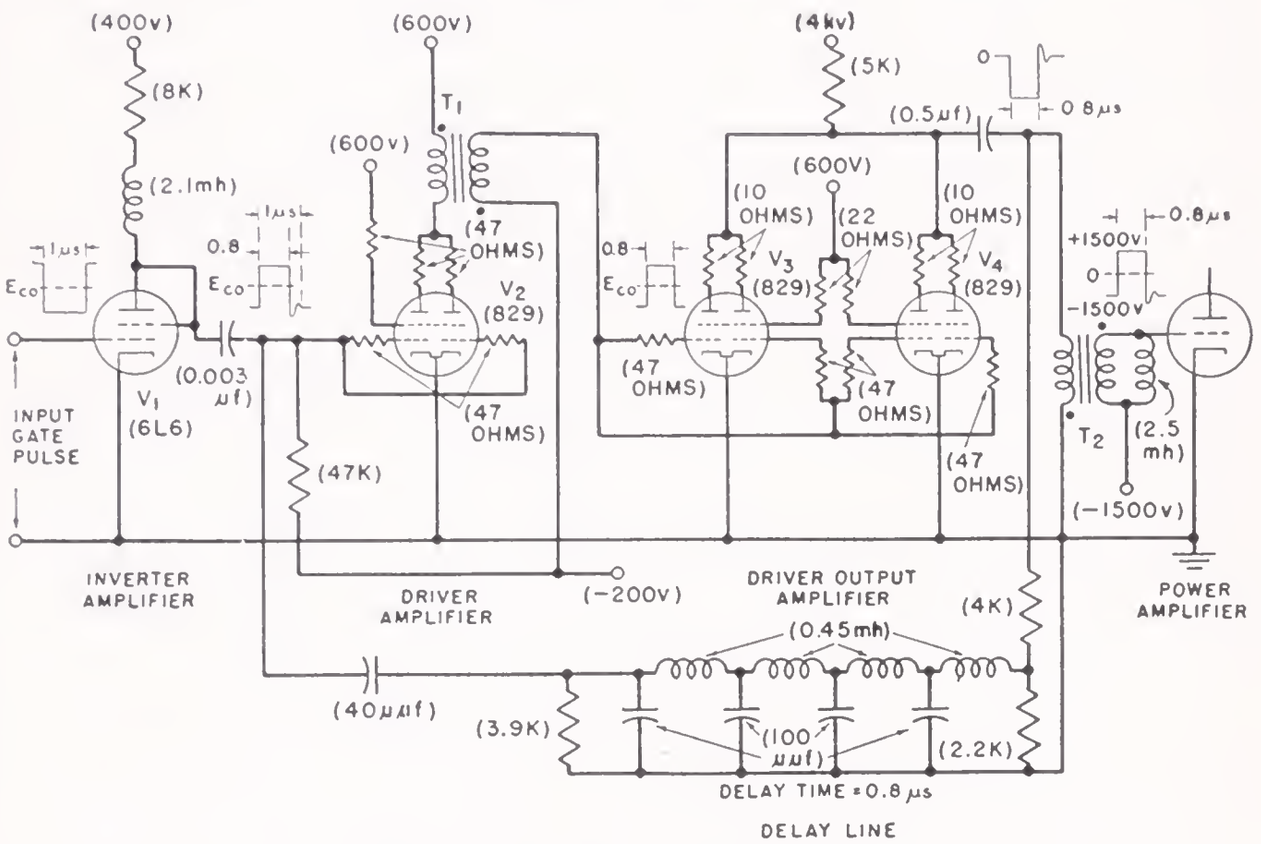


FIG. 39. Delay-line-controlled driver.

pulse ends, and therefore ringing oscillations appear in the waveforms of e_{c2}, e_{b2}, and e_o. Because of the 2-K damping resistors shunted across two of the transformer windings, these oscillations die away quickly.

14. Delay-line-controlled Driver. Driver-power-amplifier modulators capable of supplying peak powers in the range of 200 kw to 1 Mw employ several large power-amplifier tubes (such as Type 6C21 triodes) connected in parallel. The grid power required by such amplifiers is greater than that which can be produced by a bootstrap or blocking-oscillator driver using a single Type 829 amplifier tube. A driver capable of supplying the required power is shown in Fig. 39 and employs two 829 tubes in parallel as the driver-output amplifier.

Between pulses, tube V_1 in Fig. 39 is conducting, and tubes V_2, V_3, and V_4 are cutoff. At the time the transmitter is to be pulsed, a negative

gate pulse of about 1-μs duration is applied to the grid of V_1 from a timing multivibrator, cutting off that tube. The positive pulse produced at the plate of V_1 is coupled to the grids of V_2 and causes a negative pulse at the plate of V_2, a positive pulse at the grids of V_3 and V_4, a negative pulse at the primary of T_2, and a positive pulse at the grids of the power-amplifier tube. The output pulse is of shorter duration than the pulse from the multivibrator because the voltage at the primary of T_2 is fed back through a voltage divider and an 0.8-μs delay line to the grid of V_1. The negative-going front edge of the delayed pulse from the line cuts off tube V_2 before the end of the multivibrator pulse occurs.

To ensure a rectangular shape for the output pulse, the leading edge of the multivibrator pulse must be steep, because this edge controls both edges of the output pulse. The multivibrator must maintain V_1 cutoff for at least 0.8 μs in order that the output pulse may have a flat top. Tube V_1 must return to full conduction shortly after the end of the 0.8-μs pulse, because the delay line is coupled to the grids of V_2 through a short-time-constant circuit, and the delay-line pulse can maintain the tube cutoff for only a fraction of a microsecond.

The positive-going wavefront applied to the V_2 end of the delay line at the beginning of the pulse does not affect the operation of the circuit. The wavefront is transmitted to the T_2 end of the line, where negligible reflection occurs because the line is terminated in approximately its characteristic resistance. At the end of the pulse a positive-going wavefront is applied to the T_2 end of the line. This wavefront reaches the grid of V_2 0.8 μs after the end of the pulse and tends to cause a second conduction of V_2. Because the resistance of tube V_1 when conducting is much less than the characteristic resistance of the delay line, the increase of the grid voltage of V_2 is small and does not cause conduction.

A 2.5-mh inductor is connected across the output winding of T_2. Without it, the trailing edge of the power-amplifier grid-voltage pulse would not be very steep. With it, the rate of discharge of shunt capacitances is increased because of current built up in the inductance during the pulse.

15. Driver Using Shorted-end Line. In the driver circuit of Fig. 40, a pulse-forming network that simulates a *shorted*-end line is used in such a way that the driver amplifier can be cutoff during most of the repetition period and yet produce a positive pulse at its plate. Thus neither cathode coupling nor a pulse transformer is required. Furthermore, the amplitude of the voltage pulse is not limited to approximately half the supply voltage, as it would be if the pulse were formed by discharge of an open-end line. The driver circuit shown produces pulses of about the same voltage and current amplitudes as do the bootstrap and blocking oscillator driver.

Waveforms for the driver of Fig. 40 appear in Fig. 41*B*. During most

of the interval between pulses, V_1 is cutoff because its grids are returned to -150 volts. For a period of about 5 μs preceding the pulse, a positive voltage is applied to the grids, causing V_1 to conduct, and during this period a current of about 2 amp is built up in the shorted-end line in preparation for the pulse. To produce the pulse, current in V_1 is interrupted very suddenly, with the result that a positive rectangular pulse appears at the plates of V_1 and is coupled through C_c to the grids of the power-amplifier tubes. The circuit functions much like an R-L-C peaker, except that use of the pulse-forming network permits a nearly rectangular

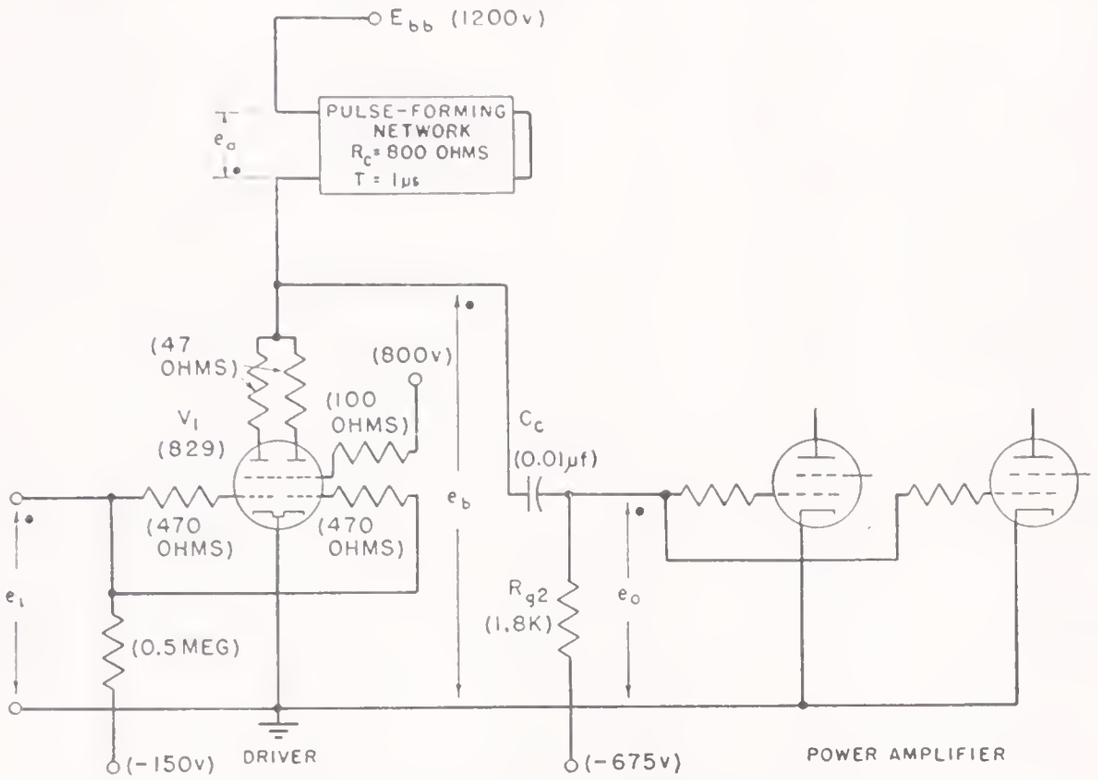


FIG. 40. Driver using shorted-end line.

pulse to be obtained instead of the more rounded pulse characteristic of a resonant circuit.

The process of starting current in the pulse-forming network is complicated by the changes of e_i during the period V_1 is conducting. Thus the i_b, e_b characteristic of the tube is variable. Nevertheless, the process is roughly the same as if the network were connected to the plate supply voltage through a battery and resistance representing the parallel combination of V_1 and the C_c, R_{g2} circuit. Current increases in several steps, as it would for an external-circuit resistance less than the characteristic resistance of the network (see Art. 3), and reaches an essentially steady-state value of 2 amp by the end of the 5-μs period. The corresponding voltage waveforms are indicated in Fig. 41*B*.

When tube V_1 is cutoff, the pulse-forming network may be represented by a resistance $R_c = 800$ ohms shunted by a 2-amp current source. The Thévenin-theorem equivalent of this parallel combination is an 800-ohm

resistor in series with a 1,600-volt battery, as in Fig. 41*A*. Capacitor C_c is charged very nearly to 1,875 volts in the period before V_1 begins to conduct. Because C_c is large, it may be considered an 1,875-volt battery during the periods of current growth and pulse production. Cutoff of

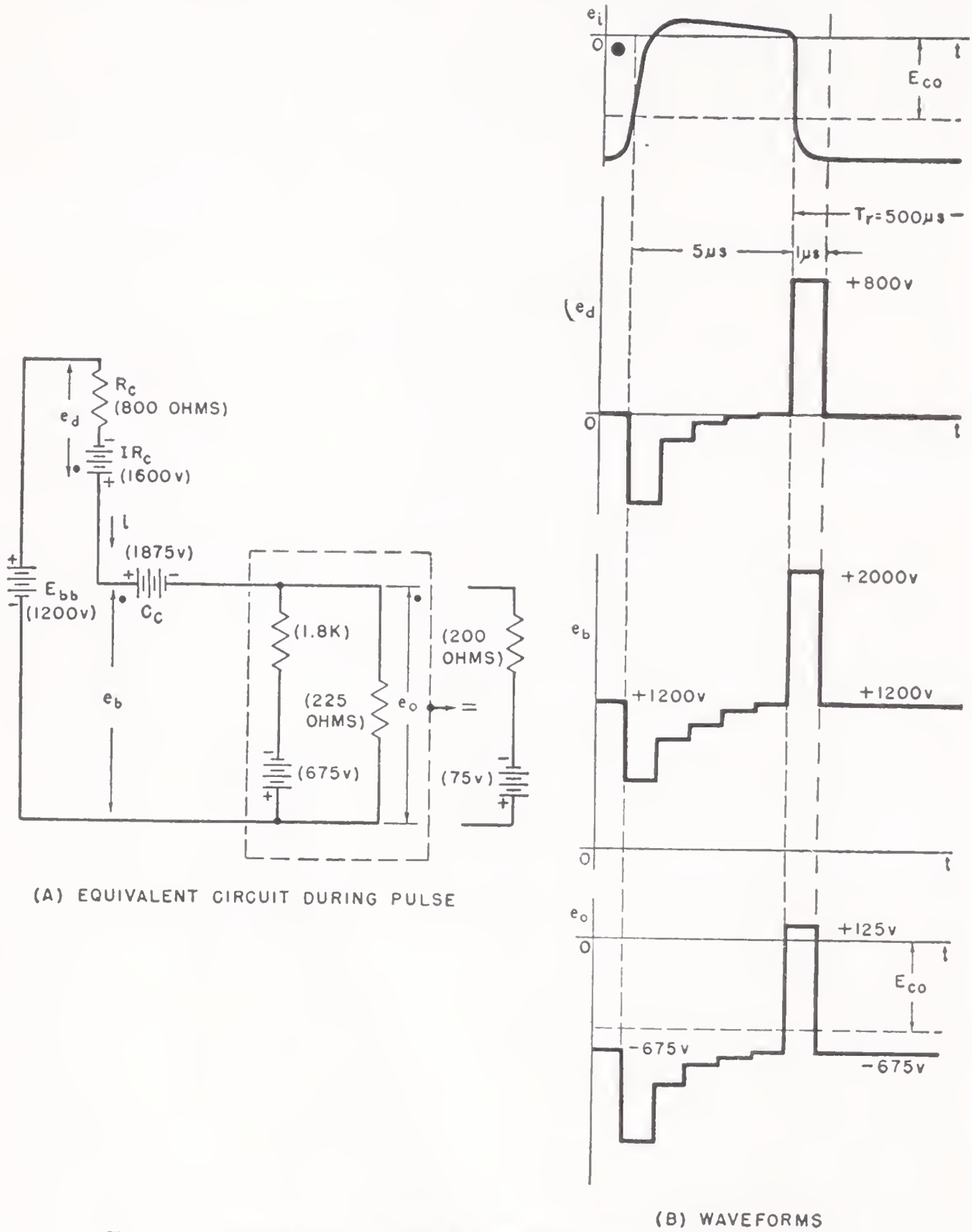


FIG. 41. Equivalent circuits and waveforms for driver of Fig. 40.

current in V_1 causes a positive voltage e_d to appear, and e_b rises above 1,200 volts and causes the power-amplifier grid voltage e_o to become positive. The entire equivalent circuit is then as drawn in Fig. 41*A*. A 225-ohm resistor is used to represent the grid resistance of the power-

amplifier tubes, usually two Type 5D21 tetrodes in parallel.[1] This equivalent circuit permits calculation of the amplitude of the pulse produced, the calculation being simplified if the portion of the circuit within the dotted rectangle is first replaced by an equivalent battery and resistor, as indicated. The network current i is found to be 1 amp, and therefore

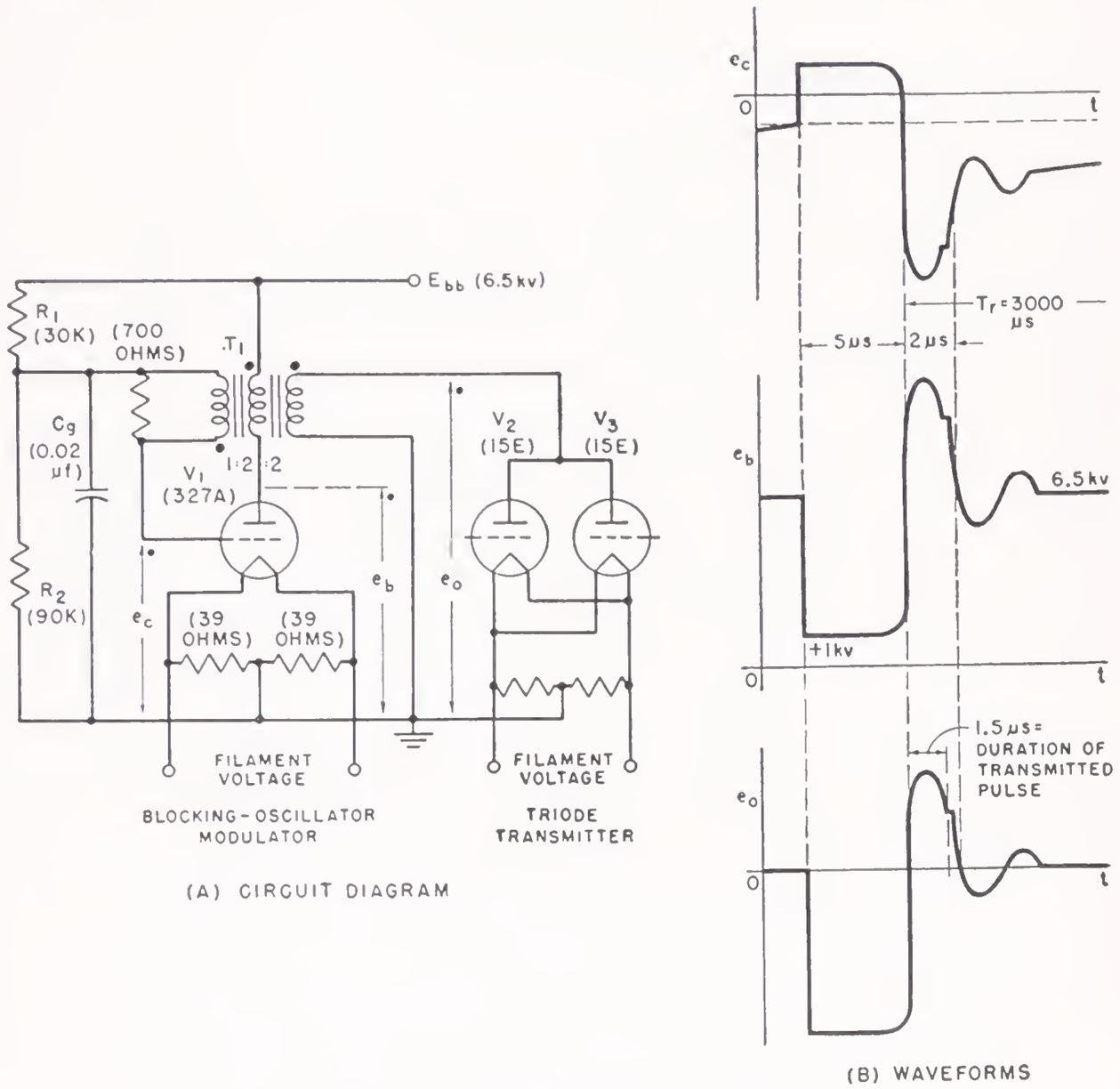


FIG. 42. Blocking-oscillator modulator.

e_o = 125 volts, e_b = 2,000 volts, and e_d = 800 volts. Since the current during the pulse is exactly half of the initial current, the effect obtained is as if the network were connected to its characteristic resistance. All the energy stored in the network is dissipated during the 1-μs pulse, and after the pulse the network is equivalent to a line with zero voltage and zero

[1] In Art. 13 the 715B (which is identical with the 5D21 except for maximum voltage rating) was assumed to have a static grid-resistance of 250 ohms when the grid voltage is 200 to 300 volts. The grid resistance of these tubes varies considerably with grid voltage and is approximately doubled when the grid voltage is reduced to 125 volts, as in the present example.

current at each point along its length. Therefore, after 1 μs, e_b returns to +1,200 volts and e_o returns to −675 volts.

16. Blocking-oscillator Modulator. In low-power radar systems, a single tube may be used as a modulator. In the circuit of Fig. 42*A*, for example, tube V_1 functions as a free-running blocking oscillator, and the pulses generated in it are applied directly to the triode transmitter tubes which produce a peak power output of about 8 kw. The repetition frequency of the radar system is controlled by the discharge of grid capacitor C_g through resistors R_1 and R_2.

The output winding of the pulse transformer in Fig. 42*A* is connected so that, when V_1 conducts, a negative pulse is applied to the plates of the transmitter tubes and no transmitter current flows (see Fig. 42*B*). During the inductive overshoot that occurs when V_1 ceases to conduct, a positive pulse is applied to the transmitter and r-f power is generated. The duration of the positive plate-voltage pulse for the transmitter depends largely upon the transformer T_1 and upon circuit capacitance. The duration of the r-f output pulse is usually shorter than the plate-voltage pulse, as indicated in the figure, because of the grid-leak bias produced in the transmitter circuit (see Art. 11, Chap. X). The shape of the output pulse is appreciably nonrectangular.

The reason for using the overswing instead of the main blocking-oscillator pulse to generate r-f energy is that by so doing the conduction period of V_1 can be increased and the peak plate current of V_1 reduced. Energy stored in the magnetic field of the transformer during the relatively long conduction period of V_1 is converted to r-f energy during the much shorter period of transmitter operation.

CHAPTER IV

INDICATORS

The component of a radar system that presents information about the location of reflecting objects is called the *indicator*. If the system is required merely to detect the presence of a reflecting object within a certain region of space, a simple indicator such as a bell, lamp, or telephone headset may be used. Usually, however, it is desired to display simultaneously the coordinates of a number of objects. Indicators using cathode-ray tubes are capable of such presentation and are the type most generally used in radar sets. Coordinates of the objects may be determined directly from the cathode-ray-tube screen or, when one object at a time is to be located accurately, from dials and counters that control the display. In the latter method, the cathode-ray tube serves to provide information for making the proper dial and counter settings. This chapter discusses cathode-ray-tube indicators and their associated circuits.

A. CATHODE-RAY TUBES

A cathode-ray tube consists of an electron gun that projects a narrow beam of electrons along the axis of the tube, a means of deflecting this beam, and a fluorescent screen to indicate the deflection of the beam. Cathode-ray tubes are usually classified according to the method used to focus and deflect the beam and are of two types. One, the so-called *electrostatic cathode-ray tube*, uses electric fields for focusing and deflecting the beam. The other employs magnetic fields for both purposes and is called a *magnetic cathode-ray tube*.

Magnetic cathode-ray tubes are capable of producing well-focused beams of high current density and are operated with higher accelerating voltages than electrostatic tubes. Because of these characteristics, an extremely bright spot may be obtained on the screen of magnetic tubes. These tubes are particularly well suited for presentations such as Types B and P, where the antenna is rotated to scan a large region. The time during which any given area of the screen is bombarded is a very small fraction of the time of the scanning cycle, and the brightness of the indicator spot is low unless the beam contains a large number of high-velocity electrons. Magnetic tubes also have the advantages of being more rugged than electrostatic tubes and of being shorter for a given screen diameter.

In applications where weight is important, electrostatic tubes have the advantage of not requiring heavy coils for deflection and focusing. The circuits used to provide sweep voltages linear with time are generally simpler than circuits required with magnetic tubes, especially when fast sweeps are needed. Electrostatic tubes are therefore used in deflection-modulation indicators where a high screen illumination is not required and in intensity-modulated indicators where weight is important and fast linear sweeps are necessary.

1. Electrostatic Cathode-ray Tubes. A schematic diagram of a typical electrostatic cathode-ray tube is shown in Fig. 1. The electron

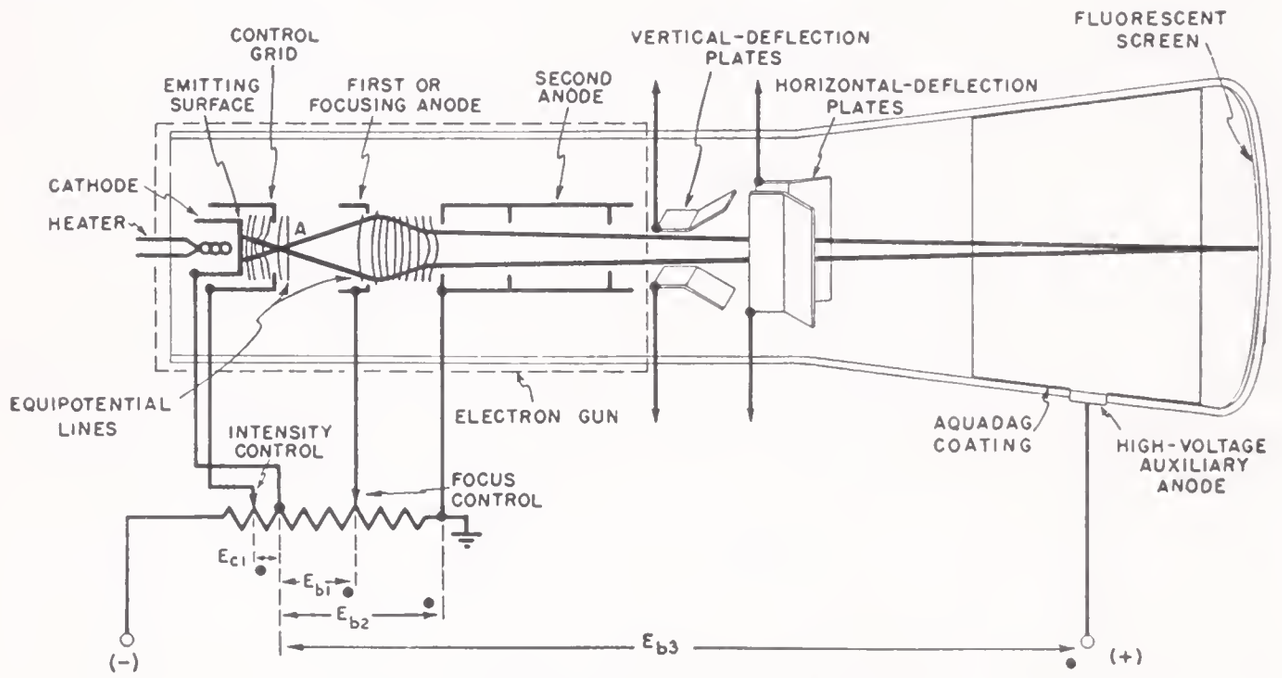


FIG. 1. Electrostatic cathode-ray tube.

gun uses electric fields to form a beam by accelerating and focusing electrons from the cathode, and the beam is deflected by electric fields between each pair of deflecting plates.

Electron Gun. The electrodes contained within the dashed-line rectangle of Fig. 1 comprise the electron gun. The cathode is a source of electrons, the control grid limits the number of electrons traveling toward the screen, and the first and second anodes focus and accelerate the electrons. These electrodes are cylinders, the axes of which are coincident with the axis of the tube. The flat end of the cathode is oxide-coated and heated by means of a noninductively wound filament. Electrons leave the cathode in random directions, but most of them are converged toward the axis of the tube by the action of the electric field between cathode and control grid. Electrons that converge sufficiently pass through an aperture in the grid and follow paths that meet at point *A*, Fig. 1, near the opening.

Convergence at point *A* is the result of the electric field produced by the combined effects of the cathode, control-grid, and first-anode voltages

on the motions of electrons traveling toward the opening. The field in this region is described by the equipotential lines shown in Fig. 1. These lines are constant-voltage contours. The forces exerted by the electric field upon the electrons accelerate them in directions perpendicular to the equipotential lines. After the electrons leave point A, their paths diverge as they approach the first anode.

The electrons pass through a space within the first anode that is almost field-free and enter the space between the first and second anodes, where the field indicated by the equipotential lines in Fig. 1 exists. Electrons not on the tube axis are deflected toward the axis as they pass through this field because they are accelerated by forces perpendicular to the equipotential lines. As the electrons enter the second-anode region, they are accelerated away from the axis, and their paths become less rapidly convergent. They have still, however, some velocity toward the axis and will meet at a point on the axis. If this point is at the screen, the beam is said to be in *focus*. The point of convergence depends upon the ratio of the voltages applied to the first and second anodes and may be changed by adjusting either of these two voltages. Focus is usually controlled by adjusting the first-anode voltage—hence the name *focusing anode*. The baffle plates shown inside the first and second anode cylinders serve to intercept electrons that follow paths too divergent for sharp focusing.

Beam Deflection. The electron beam is deflected by electric fields produced by voltages applied to two pairs of plates located just beyond the second anode. The two sets of plates are mounted at right angles to one another, the plane of one set being horizontal and the plane of the other set vertical. If a voltage is applied to the plates of a pair, the electron beam is attracted toward the more positive plate; the amount the beam is deflected is directly proportional to the voltage between the plates. As shown in Fig. 1, the plates are made parallel for a small part of their length and divergent for the remainder in order to prevent the beam from striking the plates when large deflection voltages are applied.

The intensity of the spot on the cathode-ray-tube screen depends upon the velocity of the electrons at the screen and the number of electrons per second that arrive at a given area of the screen. The spot intensity is usually controlled by adjustment of the control-grid voltage. Decreasing the magnitude of a negative grid voltage increases the beam current and therefore increases the brightness of the spot. Too large beam currents result in a poorly focused spot. Intense spots of small diameter may be obtained by use of very high second-anode voltages which result in high electron velocities, but the deflection sensitivity is then decreased. Deflection sensitivity is defined as the number of millimeters deflection at the screen caused by a 1-volt signal impressed upon a pair of deflection plates. (The term *deflection factor* is also used and is defined as the

number of volts that must be impressed upon a pair of deflection plates to produce a 1-in. deflection of the beam.)

To permit the use of higher accelerating voltages (and so provide greater screen illumination) without appreciable loss in deflection sensitivity, some tubes have an auxiliary high-voltage anode located between deflection plates and screen. This electrode may be operated at a voltage 30 to 100 per cent higher than the second-anode voltage and consists of an Aquadag coating on the inside of the tube. Aquadag is a colloidal solution of graphite which, when painted on the glass of the tube and allowed to dry, forms a high-resistance coating. Since the current density in the coating is small, the high resistance is no disadvantage.

Defocusing Effects. To avoid defocusing the electron beam, the deflection plates and second anode should be at approximately the same voltage. It is also desirable from the standpoint of external circuits that the deflection plates be near ground voltage. Both requirements may be met by connecting the second anode to ground and operating the other electrodes, except the auxiliary high-voltage anode, at negative voltages with respect to ground.

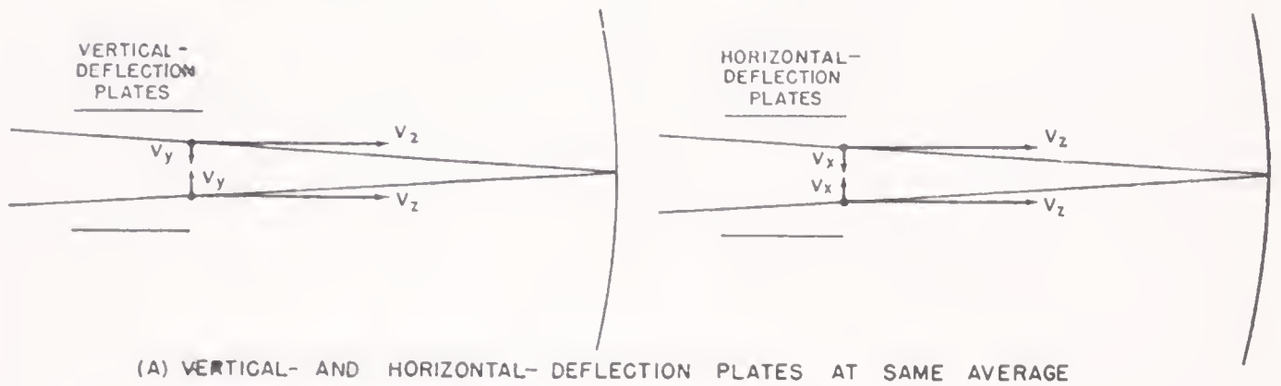
One defocusing effect in electrostatic cathode-ray tubes is called *astigmatism.* This term has the same general meaning here as in optics; it indicates that the beam is in focus in the plane of one pair of deflecting plates but out of focus in the plane of the other pair. Astigmatism results if the average voltage of one pair of deflection plates with respect to the second anode is incorrect.

Consider first that the beam is focused in both the horizontal and vertical planes with no voltage between deflecting plates, as shown in Fig. 2*A*. As the beam passes between a set of plates, the electrons in the beam have horizontal, vertical, and axial components of velocity v_x, v_y, and v_z, respectively. If, now, both vertical-deflection plates are made positive with respect to the second anode, the resultant electric field accelerates electrons toward the positive plates, thus reducing the vertical components of velocity v_y but not affecting the horizontal components v_x (compare Figs. 2*A* and *B*). The result is that the beam is defocused in the vertical plane but not in the horizontal; the point of focus in the vertical plane lies beyond the screen. A similar effect is produced if both vertical-deflection plates are made negative with respect to the second anode; the point of focus in the vertical plane then occurs between screen and deflection plates.

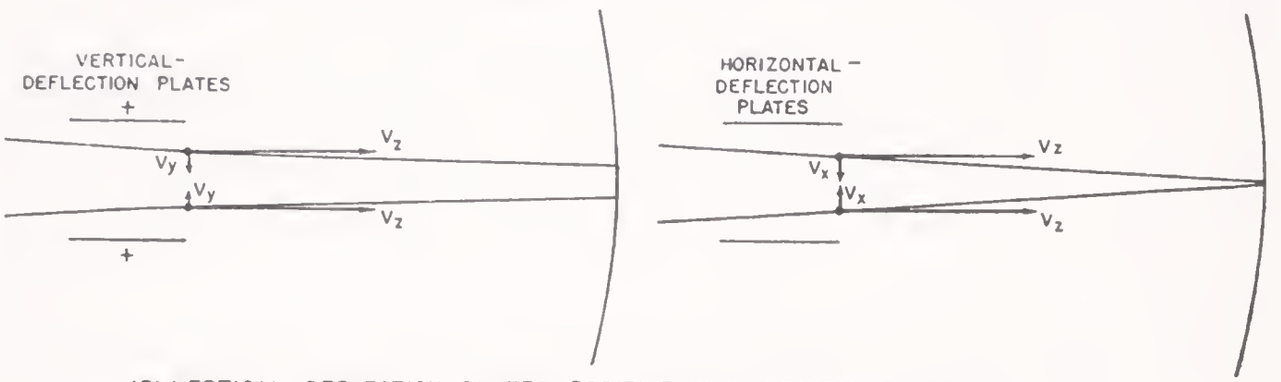
Astigmatic defocusing cannot be corrected by readjustment of the focusing-anode voltage, because no one voltage produces a focus in both the horizontal and the vertical planes. The positive voltage applied to the vertical-deflection plates in Fig. 2*B* also increases the axial velocity component v_z. Change of v_z alters the time taken by the electrons to

reach the screen and thus affects the focus in both planes. Since the effect on focus in the two planes is nearly the same, the focusing-anode voltage can be adjusted to compensate for the effect of axial velocity on focus.

To correct astigmatism, a voltage is applied between the second anode and either pair of deflection plates in such a way as to focus the beam at the same point in both planes. The point of convergence is then brought to the screen by adjustment of the main focus control. In some radar indicators an astigmatism control is provided, and the sharpest possible



(A) VERTICAL- AND HORIZONTAL- DEFLECTION PLATES AT SAME AVERAGE VOLTAGE AS SECOND ANODE



(B) VERTICAL—DEFLECTION PLATES POSITIVE WITH RESPECT TO SECOND ANODE. HORIZONTAL—DEFLECTION PLATES AT SECOND-ANODE VOLTAGE

FIG. 2. Astigmatism in cathode-ray tube.

pattern is obtained by making successive adjustments of the main-focus and the astigmatism controls.

Another defocusing effect is produced if a time-varying deflection voltage is applied to only one plate of a pair, the second plate being connected to the second anode. Variations of the deflection voltage cause the mean voltage of the pair of plates to vary with time. Variation of this mean voltage causes the point of focus in the plane of deflection to vary from point to point along the trace because of variation of transverse-velocity components and causes the point of focus in both planes to vary along the trace because of variation of axial-velocity components. Neither the main-focus control nor an astigmatism control can remove this type of defocusing—the position along the trace at which focus is obtained can be changed, but the entire trace cannot be brought into focus.

This type of defocusing is greatly reduced if the deflection voltage is applied in push-pull to the two plates of a pair. Thus the average voltage of each pair is maintained constant relative to the second anode. Push-pull sweep circuits are frequently required for this reason, but push-pull signal-deflection circuits for Type A indicators are seldom required because of the small amplitude of the signal.

Fluorescent Screens. The screen materials used in cathode-ray tubes are called phosphors. The color and persistence of the light emitted by the screen depend upon the composition of the particular phosphor used. The phosphors are designated by the letter P followed by a number. Most cathode-ray tubes of radar indicators employ a P1, P4, or P7 screen. The P1 and P4 screens have medium persistence and emit green and white light, respectively. The P7 screen has long persistence and emits yellow light.

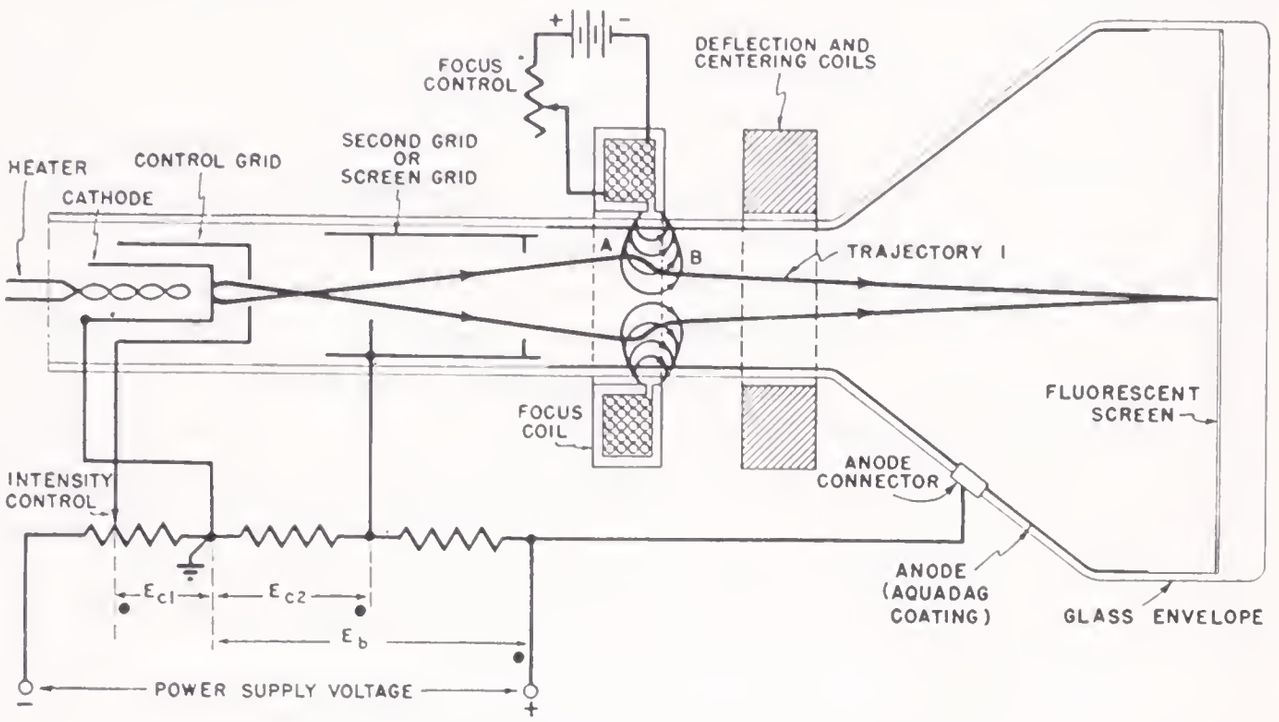
The decay time of long-persistence phosphors is greatest when they are excited at low energy densities. Because of this property, and also because the beam electrons are barely able to penetrate the phosphor, long-persistence screens cannot be excited directly by electron bombardment. The high energy density needed to produce an ample amount of light would result in short persistence. In the P7 screen, excitation at low energy densities is achieved by coating the long-persistence layer with a second layer called the flash layer. Electrons striking the flash layer cause it to emit a brilliant flash of blue light. Because the light rays are able to penetrate the long-persistence phosphor more deeply, adequate excitation is obtained at low energy densities, and a long-persistence yellow glow is produced.

The brilliant blue light given off by the flash layer would be hard on the eyes were its brilliance not reduced by placing an amber-colored filter in front of the tube. The filter also provides contrast by partially excluding ambient light. Markings to indicate the scale of coordinates are sometimes carried by the filter.

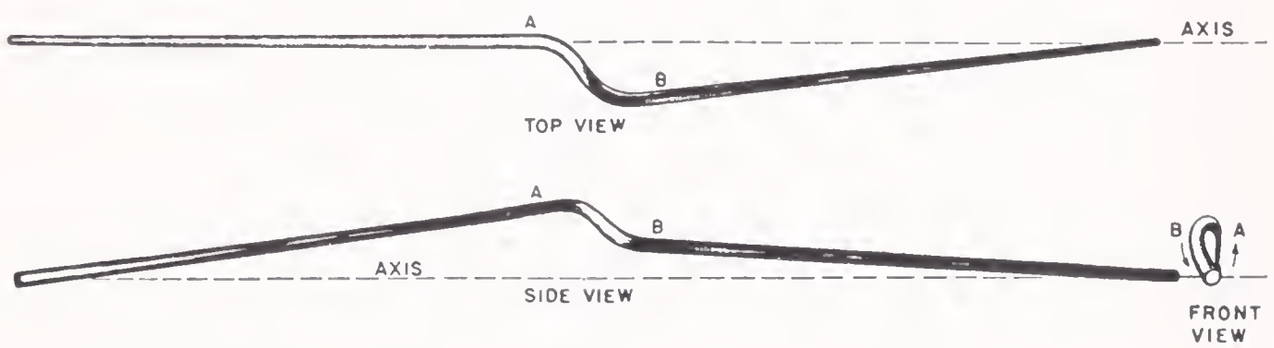
Most electrostatic cathode-ray tubes used in radar indicators employ either P1 or P4 screens; a few have P7 screens. With a P7 screen, an auxiliary anode operating at about 4 kv is necessary, because electrons must strike the screen with an energy of about 4,000 electron volts in order to produce satisfactory luminescence.

Once the electrons have reached the screen and have given up their energy in producing luminescence, they must be removed; otherwise, a large negative charge would build up on the screen. Since both the screen material and the glass of the tube are good insulators, a current of electrons cannot flow through these materials. Electrons striking the screen cause other electrons to be emitted from the screen by secondary emission. Some of these secondary electrons have sufficient velocity on emis-

sion to reach the Aquadag coating and thus to return to the power supply. Those electrons not having sufficient velocity return to the screen. An equilibrium condition is produced in which just as many secondary electrons return to the power supply as reach the screen in the beam, and the screen is thus maintained at a constant voltage. For the screens used with cathode-ray tubes of radar indicators, equilibrium is reached when



(A) MAGNETIC CATHODE-RAY TUBE CONSTRUCTION



(B) DETAIL OF ELECTRON TRAJECTORY I (NOT TO SCALE)

FIG. 3. Magnetic cathode-ray tube.

the screen is at a voltage about 100 volts below that of the collecting anode.

2. Magnetic Cathode-ray Tubes. The construction of a typical magnetic cathode-ray tube is shown in Fig. 3*A*. Note that magnetic rather than electric fields are used for focusing and for deflection. The flare angle is made greater than in an electrostatic tube to accommodate the greater angle of deflection obtainable with magnetic fields. Thus the magnetic tube is shorter for a fixed screen diameter.

Electron Gun. In the magnetic tube, the electron gun comprises the heater, cathode, control grid, second grid or screen grid, focusing magnetic field, and anode. Operation of the section of the tube that includes heater, cathode, and control grid is the same as for the corresponding parts of an electrostatic tube. The second grid serves to screen the region between cathode and control grid from the high accelerating voltage of the anode and thus reduces the value of negative control-grid voltage necessary to cut off the beam current. This electrode does not serve to focus the beam.

Focusing is accomplished by means of a coil consisting of many turns of wire wound on a soft-iron ring provided with an annular air gap, as shown in Fig. 3*A*. An adjustable direct current in the winding sets up a strong magnetic field through the center of the coil. Electrons moving exactly along the axis of the tube pass through the focusing field with no deflection, since they move parallel to the magnetic field at all times. All other electrons approach the focusing field with outward radial components of velocity. As these electrons pass through the field, they are acted on by forces which cause rotation about axes parallel to the tube axis and which produce inward radial accelerations sufficient to reverse the initial outward radial-velocity components and to bring the beam to a focus at the screen.

Consider, for example, an electron following trajectory 1 in Fig. 3*A*. At point *A*, where the electron enters the focusing field, the magnetic field has an appreciable radial component. The large *axial velocity* of the electron in the presence of this *radial* field results in a force which is in a direction to deflect the electron outward from the plane of the page. The electron therefore enters the central portion of the magnetic field with a tangential component of velocity (a component perpendicular to a radial line from the tube axis to the electron); thus electrons in the central portion of the field tend to rotate about the tube axis.

The central portion of the focusing magnetic field is almost entirely axial, and consequently in this region the axial component of the electron velocity does not cause a deflection force. However, the *tangential* component of velocity in the presence of the *axial* magnetic field results in a force which causes the electron to be accelerated *radially inward* toward the tube axis. The electron follows an approximately helical path as it passes through the region between points *A* and *B*, as indicated in Fig. 3*B*, and acquires an inward radial component of velocity.

In leaving the magnetic field at point *B*, the electron again passes through a region in which there is a radial component of magnetic field. The tangential force is now in the opposite direction, with the result that the tangential velocity imparted to the electron as it entered the field is canceled. The inward radial velocity component acquired in the region

between A and B is not canceled, however. The electron beam, therefore, rotates only while it is passing through the magnetic field. The electron emerges with the same axial velocity with which it entered the field and with a radial velocity component which is now inward and causes the electron to travel toward the axis of the tube. It may be shown that all electrons in the beam reach the axis at the same point. By adjusting the current in the focusing coil, this point may be made to lie on the screen, thereby producing a sharply focused spot. The focusing action depends upon the shape, strength, and axial position of the magnetic field. Tilting the focus coil causes the beam to tilt away from the tube axis, and therefore an adjustment that tilts the coil may be used to center the spot on the screen.

Instead of a focusing coil, a permanent magnet may be used. The magnet may have the form of a two-section annular ring, as shown in Fig. 4. The focusing effect is adjusted by changing the air-gap spacing between the pole pieces. In some applications, the magnet is moved axially to bring the trace into focus and tilted to center the pattern.

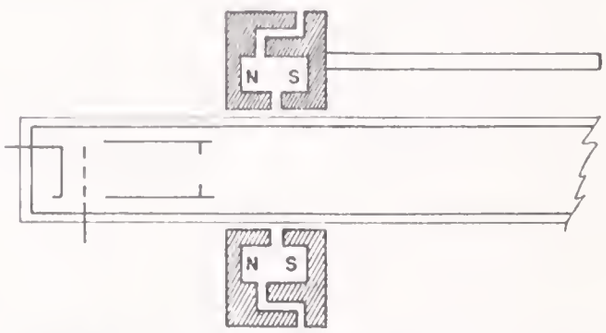


FIG. 4. Permanent magnet for focusing electron beam.

Beam Deflection. The electron beam is deflected by the magnetic field that results from current passing through a set of deflection coils. The coils are many-turn windings that may take any of several forms and may be wound on air or ferromagnetic cores. The coil assembly, called the *deflection yoke*, is mounted in a magnetic shield and placed around the neck of the cathode-ray tube. Four types of deflection coils are illustrated in Fig. 5.

In one method of obtaining plan-position indication, the beam is deflected in a fixed direction with respect to the yoke, and the yoke itself is rotated to change the direction of the deflection on the screen. The deflection yokes of Figs. 5*A*, *B*, and *C* may be used in this way. The air-core coils of Fig. 5*A* are bent to fit closely around the neck of the cathode-ray tube, and the coil current sends magnetic flux through the tube in the direction shown. The return path of the flux is through the magnetic shield that surrounds the core. To secure a more uniform distribution of flux, the turns of the coils are often distributed, as in Fig. 5*B*. Sometimes a distributed winding is used with its turns placed in the slots of an iron core. The assembly is then called a motor-stator yoke because of its appearance. A different iron-core yoke carries toroidal windings, as illustrated in Fig. 5*C*. The windings are connected so that the magnetic fluxes are in opposite directions in the core but the same direction across

the neck of the tube. The electron beam is always deflected in a direction perpendicular to the magnetic flux. Thus, for the assemblies of Figs. 5*A*, *B*, and *C*, the direction of deflection changes as the yoke is rotated.

When fixed deflection yokes are employed, two sets of windings are required to produce horizontal and vertical deflections. The yoke of Fig. 5*D*, employing a square iron core, may be used. Current in coils *A* and *B* causes a vertical flux through the neck of the cathode-ray tube and a horizontal deflection. Current in windings *C* and *D* produces a horizontal flux and vertical deflection. Because of the symmetrica construc-

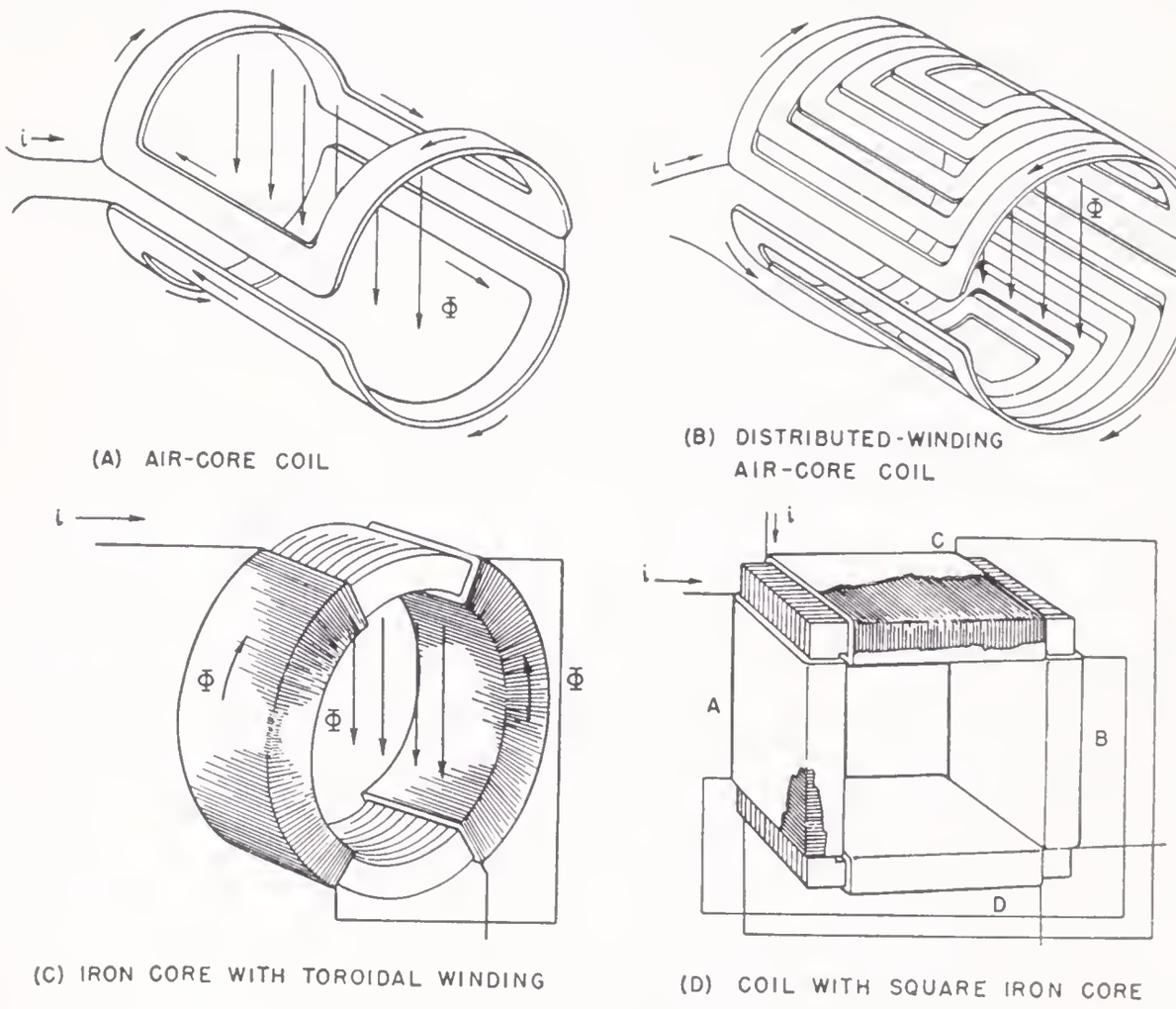


FIG. 5. Types of deflection coils.

tion, fluxes from windings *A* and *B* induce equal-magnitude voltages of opposite polarity in windings *C* and *D*. The same is true of the fluxes set up by windings *C* and *D* which link coils *A* and *B*. Fixed deflection yokes capable of providing deflections in two directions are also made in the forms of Figs. 5*A* and *B*, with a second set of windings added at right angles to the windings shown. In each structure, careful balancing of induced voltages is needed to prevent interaction between the two sets of windings.

Fluorescent Screens. The magnetic cathode-ray tubes of radar indicators use long-persistence phosphors such as the P7. A long-persistence screen is required because these tubes are normally used with presenta-

tions such as Types B and P, where the antenna scans a large region and any one object is illuminated for only a short time during the scanning cycle time.

3. Designations and Ratings of Cathode-ray Tubes. Cathode-ray tubes are designated by a numeral and a letter, followed by the screen designation. The numeral signifies the diameter of the screen in inches, and the letter the order of development. For example, the Type 5BP1 tube has a 5-in. screen of P1 phosphor. A similar tube using P4 phosphor is designated as Type 5BP4. Table 1 gives a list of cathode-ray tubes commonly used in radar together with their ratings.

Table 1. Ratings of Typical Cathode-ray Tubes Used in Radar Indicators

Type	*Focus and deflection method*	*Maximum voltage ratings**			*Typical operation†*			
		E_{b2} *or* E_b, *volts*	E_{b1} *or* E_{c2}, *volts*	E_{c1}, *volts*	E_{b2} *or* E_b, *volts*	E_{b1} *or* E_{c2}, *volts*	E_{c1} *for visual cutoff, volts*	*Deflection factor, volts/in.*
3BP1	Electrostatic	2,200	1,100	Never positive	2,000	575	−60 ± 50%	200 ± 20%
3FP7‡	Electrostatic	2,200	1,100	Never positive	2,000	575	−60 ± 50%	250 ± 20%§
5BP1	Electrostatic	2,200	1,100	Never positive	2,000	450	−40 ± 50%	84 ± 17%
5CP1‡	Electrostatic	2,200	1,100	Never positive	1,500	431	−45 ± 50%	55 ± 20%‖
5FP7 7BP7 12DP7	Magnetic	7,700	750	Never positive	7,000	250	$-45 \begin{matrix} +25\% \\ -20\% \end{matrix}$	

* All voltages are measured with respect to cathode.
† All tubes listed operate with heater voltage of 6.3 volts.
‡ Has an auxiliary high-voltage anode. Maximum voltage rating is 4,400 volts.
§ With auxiliary high-voltage anode at 4,000 volts.
‖ With auxiliary high-voltage anode at 1,500 volts.

4. Data Presentation by Means of Cathode-ray Tubes. The three basic methods of presenting radar data by means of cathode-ray tubes, described in Art. 3, Chap. I, are Type A, Type B, and plan-position (Type P) indication. In Type A indication, only the range of the object detected is indicated on the screen, but in Types B and P, the two coordinates available on the screen of a cathode-ray tube are used to indicate both range and azimuth. To describe completely the position of a reflecting object, three coordinates are required. Which two of these three coordinates are of greatest importance depends upon the application. Therefore, many different ways of presenting data have been developed to display different pairs of coordinates in a variety of forms. Some of the displays provide rough indication of a third coordinate.

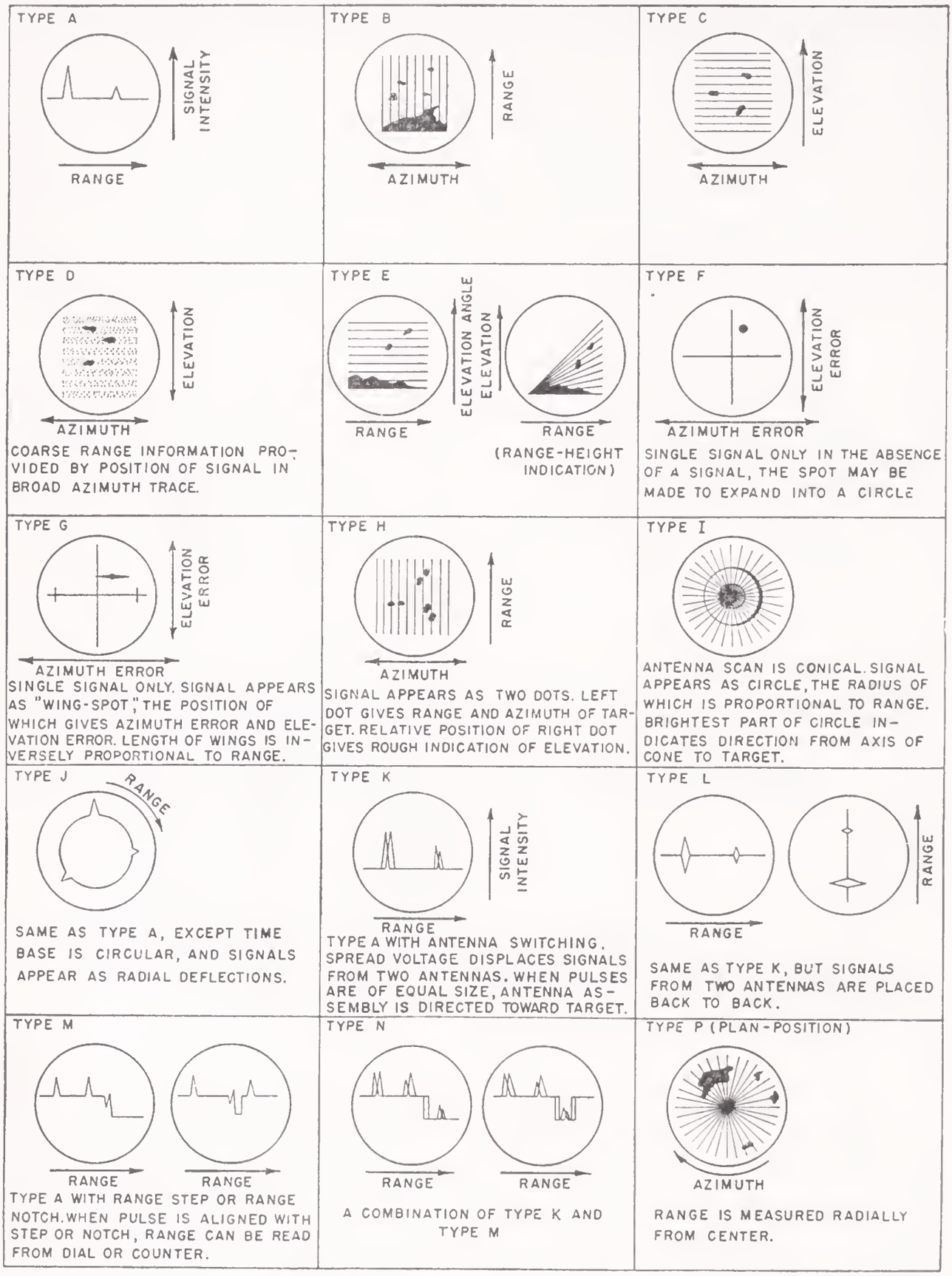

TYPE A
SIGNAL INTENSITY
RANGE
TYPE B
RANGE
AZIMUTH
TYPE C
ELEVATION
AZIMUTH
TYPE D
ELEVATION
AZIMUTH
COARSE RANGE INFORMATION PROVIDED BY POSITION OF SIGNAL IN BROAD AZIMUTH TRACE.
TYPE E
ELEVATION ANGLE
ELEVATION
RANGE
RANGE
(RANGE-HEIGHT INDICATION)
TYPE F
ELEVATION ERROR
AZIMUTH ERROR
SINGLE SIGNAL ONLY IN THE ABSENCE OF A SIGNAL, THE SPOT MAY BE MADE TO EXPAND INTO A CIRCLE
TYPE G
ELEVATION ERROR
AZIMUTH ERROR
SINGLE SIGNAL ONLY. SIGNAL APPEARS AS "WING-SPOT," THE POSITION OF WHICH GIVES AZIMUTH ERROR AND ELEVATION ERROR. LENGTH OF WINGS IS INVERSELY PROPORTIONAL TO RANGE.
TYPE H
RANGE
AZIMUTH
SIGNAL APPEARS AS TWO DOTS. LEFT DOT GIVES RANGE AND AZIMUTH OF TARGET. RELATIVE POSITION OF RIGHT DOT GIVES ROUGH INDICATION OF ELEVATION.
TYPE I
ANTENNA SCAN IS CONICAL. SIGNAL APPEARS AS CIRCLE, THE RADIUS OF WHICH IS PROPORTIONAL TO RANGE. BRIGHTEST PART OF CIRCLE INDICATES DIRECTION FROM AXIS OF CONE TO TARGET.
TYPE J
RANGE
SAME AS TYPE A, EXCEPT TIME BASE IS CIRCULAR, AND SIGNALS APPEAR AS RADIAL DEFLECTIONS.
TYPE K
SIGNAL INTENSITY
RANGE
TYPE A WITH ANTENNA SWITCHING. SPREAD VOLTAGE DISPLACES SIGNALS FROM TWO ANTENNAS. WHEN PULSES ARE OF EQUAL SIZE, ANTENNA ASSEMBLY IS DIRECTED TOWARD TARGET.
TYPE L
RANGE
RANGE
SAME AS TYPE K, BUT SIGNALS FROM TWO ANTENNAS ARE PLACED BACK TO BACK.
TYPE M
RANGE
RANGE
TYPE A WITH RANGE STEP OR RANGE NOTCH. WHEN PULSE IS ALIGNED WITH STEP OR NOTCH, RANGE CAN BE READ FROM DIAL OR COUNTER.
TYPE N
RANGE
RANGE
A COMBINATION OF TYPE K AND TYPE M
TYPE P (PLAN-POSITION)
AZIMUTH
RANGE IS MEASURED RADIALLY FROM CENTER.

FIG. 6. Types of data presentation.

Sketches of a number of indicator presentations are drawn in Fig. 6. The Types J, K, L, M, and N presentations are modification of the basic Type A display, and the Types C, D, E, H, and I are modifications of the basic Type B and Type P displays. The first group employs deflection modulation of the indicator beam and has the advantage that the circuits

required are generally simpler than those required for the second group in which intensity modulation is employed. Additional advantages of the presentations employing deflection modulation are (1) their adaptability to accurate methods of measuring range and angle and (2) their superiority in making echo signals visible in the presence of strong interference. The displays employing intensity modulation have the great advantage of presenting data in more convenient and more easily interpreted form.

The two forms of Type E presentation may be compared with the Type B and Type P presentations of range and azimuth. The form in the left-hand diagram compares with Type B—vertical displacement is proportional to elevation *angle*. Therefore, a constant-altitude line in the display appears to be lower at increased ranges. This distortion is avoided in the form of Type E presentation shown in the right-hand sketch, where the range trace is turned through an angle dependent on the elevation angle in the same way that the range trace is turned through the azimuth angle in plan-position indication. The vertical coordinate for this display is elevation *distance*, and therefore the name *range-height indication* is often applied.

B. SWEEP-VOLTAGE CIRCUITS

Electrostatic cathode-ray tubes require that varying voltages be applied to the deflection plates to cause motion of the luminescent spot on the screen. Most electrostatic tubes are used in deflection-modulated indicators; the signal voltage from the receiver is applied to one pair of deflection plates, and a *range-sweep voltage* is applied to the other pair of plates to move the spot at a constant rate during the interval of echo reception. Circuits for the generation of range-sweep voltages are the subject of the following articles.

5. Basic Type A Indicator. The diagram of Fig. 7 is an example of an elementary Type A indicator circuit. The cathode-coupled multivibrator (Art. 15, Chap. II) is triggered at the time of the transmitted pulse and generates (1) a negative gate pulse that serves to cut off the saw-tooth-generator tube for a time corresponding to the maximum range desired and (2) a positive pulse that is applied to the grid of the cathode-ray tube as an intensifying pulse. The saw-tooth generator (Art. 5, Chap. II) provides a nearly linear rising voltage which is then amplified, inverted, and applied to the left horizontal-deflection plate to produce a deflection of the beam to the right. Positive video signals at the receiver output are applied to the top vertical-deflection plate and cause upward deflections to appear along the sweep trace at distances from the transmitted pulse proportional to the ranges of the targets. Provision is made for positioning the pattern by adjusting the direct-voltage component of both the sweep and the signal waveforms. Clamper tubes (Art. 10,

Chap. II) keep the pattern in the desired position despite changes in the average value of the sweep or signal voltages.

The sweep amplifier in the circuit of Fig. 7 is necessary in order to obtain a reasonably linear sweep voltage of sufficient amplitude to deflect the beam across the tube. To obtain a very nearly linear rise of voltage from the saw-tooth generator, the time constant R_LC must be made large so that the amplitude of the saw tooth is only a small fraction of the plate supply voltage. Such a low-amplitude saw tooth must be amplified before it can be used to deflect the beam of the cathode-ray tube. For example, suppose in Fig. 7 that E_{bb} is 400 volts. A typical value of deflection factor for a 5-in. cathode-ray tube is 60 volts per in., so that, if 80 per cent of the screen diameter is used, the required sweep-voltage

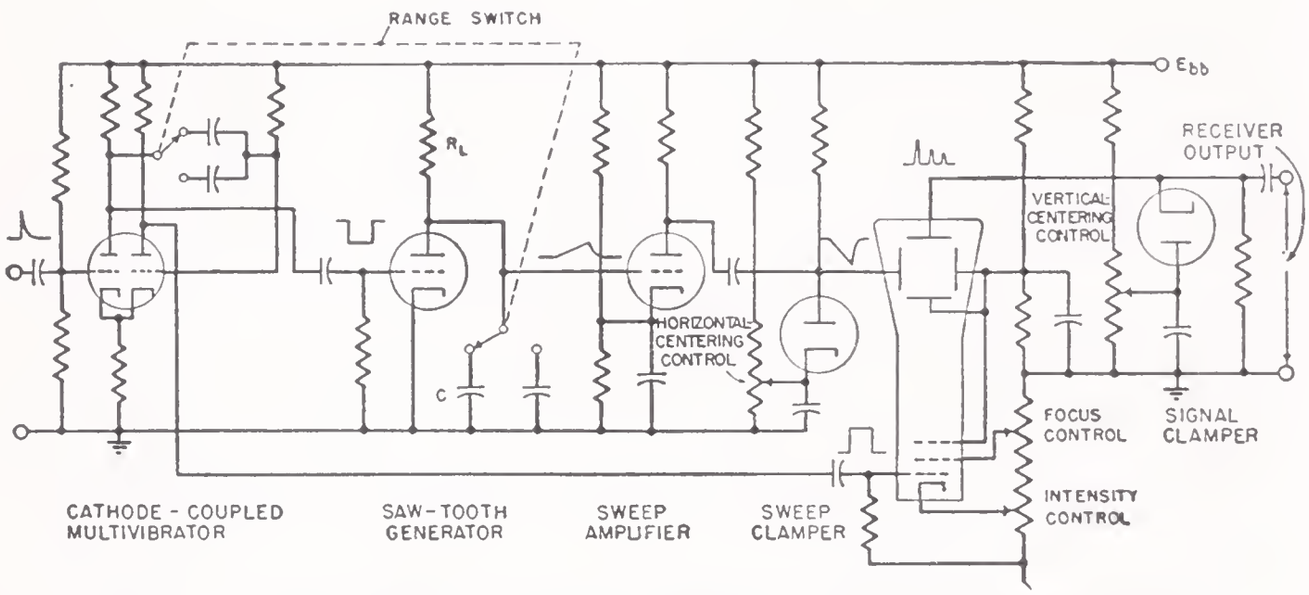


FIG. 7. Circuit diagram illustrating principles of Type A indicator circuit.

amplitude is about 240 volts for full-scale deflection. If the amplification of the sweep amplifier is 12, the voltage change required across C, the sweep capacitor, is 20 volts. Since 20 volts is only 5 per cent of E_{bb}, the capacitor charging curve is nearly linear, and the time constant R_LC must be about 20 times the sweep duration. The nonlinearity of the sweep provided by such a circuit results mainly from distortion in the amplifier rather than from the exponential nature of the sweep-capacitor charging transient.

The need of a sweep amplifier may be eliminated by using a very high plate supply voltage (say several thousand volts) to generate the 240-volt saw tooth. This method has the disadvantage of requiring high-voltage power supplies and high-voltage insulation for parts of the sweep circuit. More frequently, the special circuits described in Art. 6 are employed to produce a linear sweep voltage. Other modifications of the basic circuit of Fig. 7 include the use of a push-pull rather than a single-ended output circuit to reduce the defocusing of the beam (Art. 1, p. 220). Also, in

some circuits a pulse from the sweep-generator circuit is fed back to the multivibrator to terminate the negative gate pulse when the sweep voltage reaches a specified value. This feedback eliminates the need for changing capacitors in the multivibrator circuit when the range scale is changed.

6. Methods of Improving the Linearity of Sweep Voltages. Generating a sweep voltage of better linearity than that produced by the circuit of

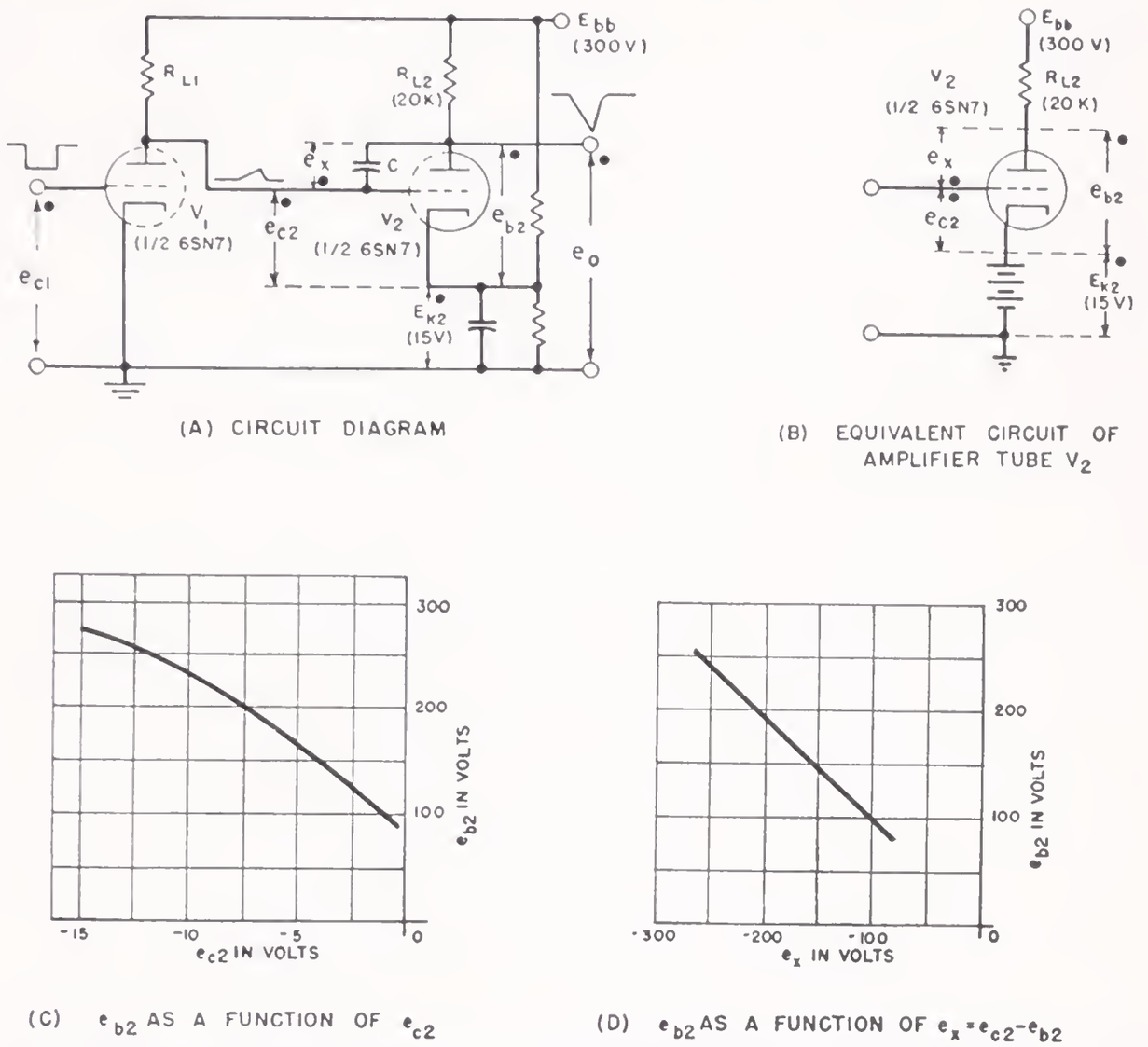


FIG. 8. Use of grid-to-plate capacitor connection to improve linearity of sweep amplifier.

Fig. 7 requires that distortion resulting from nonlinear tube characteristics as well as curvature caused by an exponential charging curve be avoided. Several circuits that are widely used in radar indicators introduce little distortion of either form and yet do not require an especially high plate supply voltage.

Sweep Capacitor Connected between Grid and Plate. If the sweep capacitor C in Fig. 7 is moved to the position shown in Fig. 8*A*, the sweep-voltage distortion caused by amplifier-tube nonlinearities is greatly reduced. To gain an understanding of why the connection improves the sweep linearity, consider first the amplifier portion of the circuit, as shown in Fig. 8*B*.

The amplifier plate-to-cathode voltage e_{b2} for various values of grid-to-cathode voltage e_{c2} may be found from a 20-K load line drawn from 285 volts on the plate characteristics of the Type 6SN7 tube. The values obtained are plotted in Fig. 8*C*. If the sweep voltage at the grid is exactly linear, then the waveform of e_{b2} during the sweep is a replica of the curve in Fig. 8*C*. This curve, therefore, shows the distortion of the sweep produced by the amplifier when the sweep capacitor is connected between grid and ground, as in Fig. 7.

The curve of e_{b2}, as a function of the grid-to-plate voltage $e_x = e_{c2} - e_{b2}$, can also be plotted for the amplifier of Fig. 8*B* by means of the plate characteristics. This curve is shown in Fig. 8*D*. Note that it is a much more nearly linear curve than that of Fig. 8*C*. For example, with a 150-volt change in e_{b2} the curve departs from linearity by about 1 volt. Consequently, if e_x can be made to vary linearly with time, a nearly linear sweep will be produced at the plate of V_2.

An almost linear variation in e_x is obtained by connecting the sweep capacitor between the grid and the plate of V_2. The change in e_{c2} required to produce a given amplitude of sweep voltage at the plate of V_2 is the same whether the sweep capacitor is connected between grid and ground or between grid and plate. Because this required voltage change is small compared with the plate supply voltage of V_1, the current through R_{L1}, and hence the charging current for the sweep capacitor, is essentially constant during the sweep. Therefore, with the grid-to-plate sweep-capacitor connection, e_x varies almost linearly with time, and the sweep-voltage waveform is a replica of the nearly linear curve of Fig. 8*D*. Note that e_x can change by a large amount during the sweep without an appreciable change in the charging current, because most of the change of e_x occurs at the plate rather than at the grid of V_2.

Another effect is produced by the capacitor connection of Fig. 8*A*. Because the voltage e_{b2} drops by A volts when e_{c2} rises by 1 volt (where A is the amplification of V_2), the changes of e_x are $(A + 1)$ times as great as those of e_{c2}. Therefore, for a given charging current to produce the same rate of change of e_{c2}, a capacitor between grid and ground must be $(A + 1)$ times as large as a capacitor between grid and plate. This effect has no influence upon sweep linearity—a smaller capacitor must be used between grid and plate than between grid and ground in order to maintain the correct sweep amplitude. A similar multiplication of capacitance between grid and plate occurs in any triode amplifier and is often called the *Miller effect*. The grid-to-plate interelectrode capacitance is equivalent to a larger capacitance between grid and cathode.

The effective amplification of capacitance is sometimes used to advantage in special devices where sweep durations of the order of minutes are required. The sweep capacitor is connected from the grid of the first

stage to the plate of the last stage of a cascade amplifier having an odd number of stages. When connected in this manner, a small capacitor produces the same effect as one several thousand times as large connected between grid and ground. Use of a prohibitively large grid-to-ground sweep capacitor is thus avoided.

Bootstrap Sweep Circuit. A second circuit for improving the linearity of a sweep voltage is drawn in Fig. 9A. The name bootstrap is sometimes applied to it because the rising sweep voltage across C is coupled through

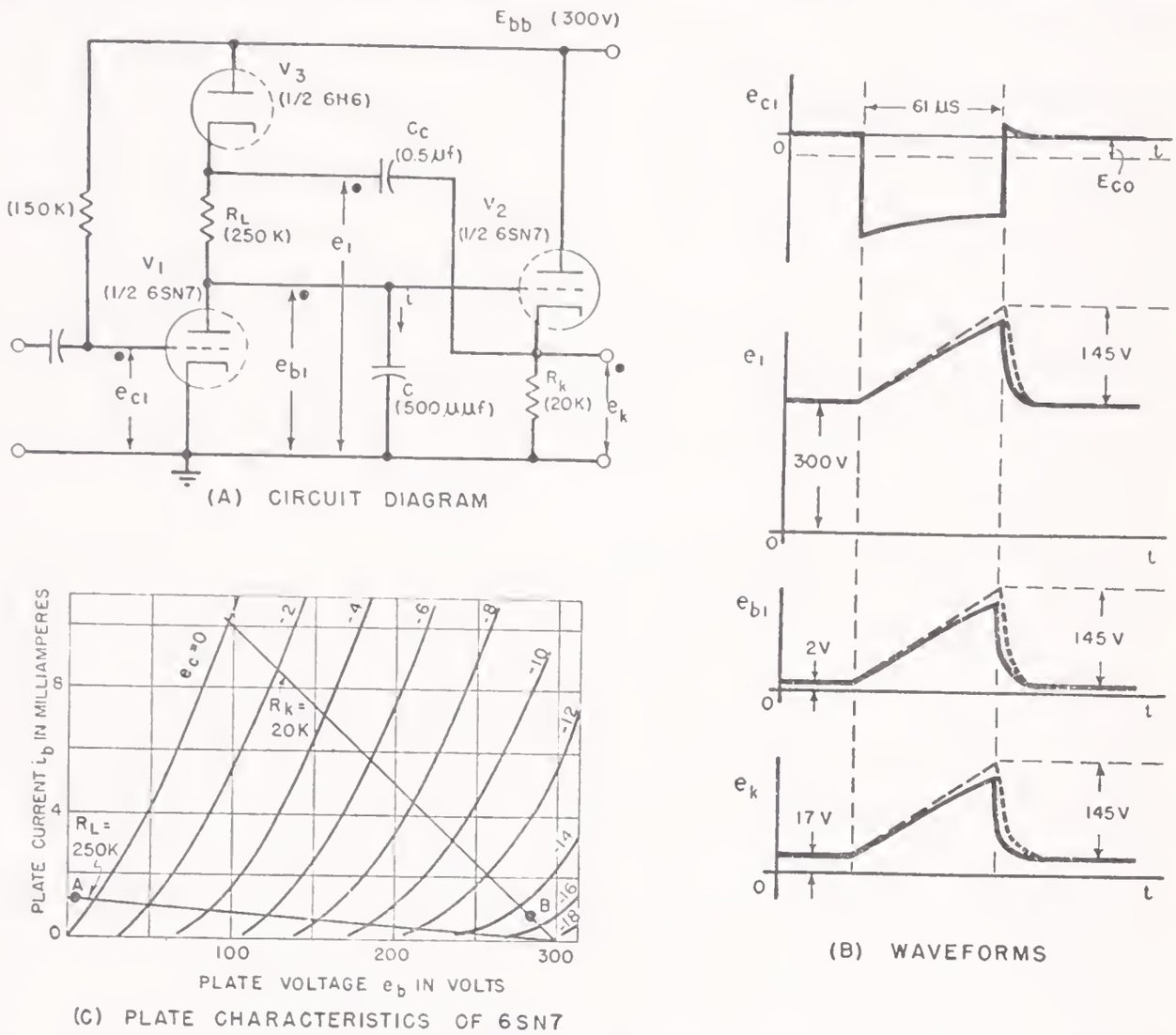


FIG. 9. Bootstrap sweep circuit.

a cathode follower (tube V_2) and a larger capacitor (C_c) to the high-voltage end of the sweep resistor (R_L). Thus, when V_1 is cut off by a negative gate pulse to produce the sweep, both ends of resistor R_L rise in voltage by nearly the same amount and the voltage across R_L remains essentially constant. As a result, the current through R_L, which is the charging current for C, is very nearly constant throughout the sweep interval; and the sweep voltage is therefore nearly linear. The diode V_3 is employed in order that the voltage applied by C_c to R_L may rise above the plate supply voltage during the sweep period. At the end of the sweep, tube V_1 conducts and discharges capacitor C. The diode main-

tains the voltage e_1 at the junction of R_L and C_c essentially equal to E_{bb} between sweeps.

Waveforms for this circuit are drawn in Fig. 9*B* for the circuit-element values indicated in Fig. 9*A* and a 10,000-yd (61-μs) sweep. The waveforms may be calculated as follows: Between sweeps the grid voltage e_{c1} of V_1 is held at approximately +1 volt by the positive-grid-return connection. The corresponding plate voltage e_{b1}, which may be determined from point A on the 250-K load line in Fig. 9*C*, is 2 volts. A trial-and-error process may be used to locate the operating point of tube V_2 on the 20-K load line (point B in Fig. 9*C*). At point B, the plate current of tube V_2 is 0.85 ma, and therefore $e_k = 0.85 \times 20 = 17$ volts. Because the voltage across diode V_3 is negligible, $e_1 = 300$ volts.

When V_1 is cut off, the initial value of the charging current for the sweep capacitor is $i = (300 - 2)/250 = 1.19$ ma, and the initial rate of rise of e_{b1} is therefore $i/C = (1.19 \times 10^{-3})/0.0005 = 2.38$ volts per μs. If the amplification of the cathode follower were unity, and if the voltage across C_c did not change during the sweep, this rate of change would be maintained throughout the 61-μs sweep interval. Thus e_{b1}, e_k, and e_1 would each rise by $61 \times 2.38 = 145$ volts, as indicated by the dashed curves in Fig. 9*B*.

Actually, the amplification of the cathode follower is less than unity, and the voltage across C_c decreases because it discharges slightly during the sweep. Both of these effects tend to decrease the voltage across R_L at the end of the sweep. Therefore, the slope of the sweep saw tooth decreases toward the end of the rise, as indicated by the solid curves in Fig. 9*B*. The curvature is in the same direction as that obtained in the simple sweep circuit with the charging resistance returned to E_{bb}, but is less in amount.

The curvature of the sweep waveform resulting from a cathode-follower amplification A is the same as that which would result if the sweep capacitor were charged through a resistance equal to $R_L/(1 - A)$. Because A is less than unity, voltage changes at each end of resistor R_L are unequal. An increase of e_{b1} by an amount Δe_{b1} causes e_1 to change by an amount $A\Delta e_{b1}$; therefore, the change in current through R_L is

$$\Delta i = \frac{\Delta e_{b1} - A\Delta e_{b1}}{R_L}$$

This equation may be written

$$\Delta i = \frac{\Delta e_{b1}}{R_L/(1 - A)}$$

which form indicates that the effective incremental resistance of the charging-current path is multiplied by the factor $1/(1 - A)$.

The advantage of the bootstrap circuit over the conventional saw-tooth-generator circuit is that the initial slope of the sweep-voltage waveform is determined by the relatively low charging resistance R_L while the curvature of the sweep depends upon the much higher value $R_L/(1 - A)$. Therefore, a larger sweep amplitude can be obtained with a given plate supply voltage without excessive nonlinearity.

Sweep Integrating Circuit. A further improvement of sweep linearity can be obtained if the sweep capacitor in Fig. 9*A* is divided into two parts and a feedback resistor R_f added, as in Fig. 10. The rising voltage e_k is applied to the integrating circuit composed of R_f and C_2. The resultant rising waveform at point *A* is concave upward, because capacitor C_2 is charged at an increasing rate as the sweep progresses. The concave-upward voltage across C_2 adds to the voltage across C_1, which is concave downward because of the decrease of current in R_L during the sweep. If the value of R_f is correctly chosen, the two curvatures very nearly cancel in the sum of the two voltages, and the output voltage waveform is extremely close to linear.

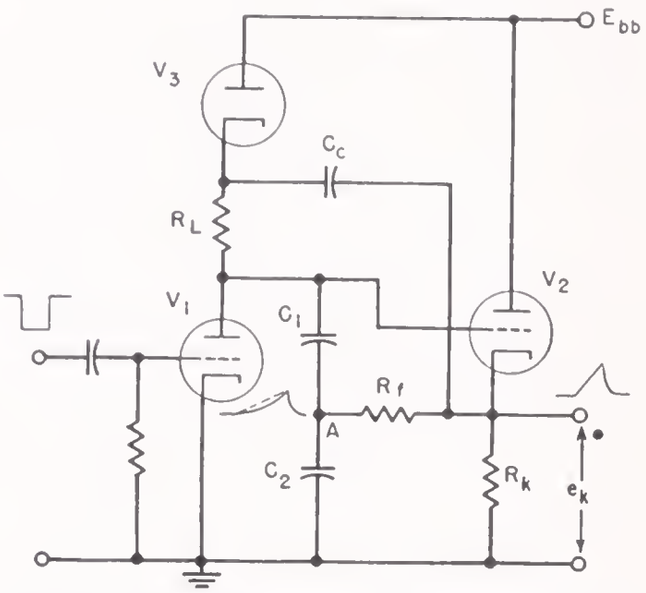


FIG. 10. Integrating circuit added to bootstrap sweep generator.

Where the circuit of Fig. 10 is required to operate at different repetition rates to provide several different sweep lengths, a diode clamper tube is often placed in parallel with the feedback resistor to discharge the integrating capacitor quickly after the sweep. This ensures that the integrating-capacitor voltage at the beginning of the sweep is always the same, regardless of the maximum range or pulse repetition rate used.

Constant-current Tubes. Sometimes a tube is used instead of a resistor to charge or discharge a capacitor and produce a saw-tooth waveform. The saw tooth thus produced is essentially linear because the tube is operated so that its plate current is maintained constant, despite the changes of plate voltage that occur during the sweep. One example of a constant-current tube is a pentode operated above the knee of the plate characteristics where the current is nearly independent of the plate voltage. A second example is either a triode or a pentode with its grid connected to the negative end of a large cathode resistor. Such a tube draws an essentially constant current, because any tendency of the plate current to change alters the grid voltage in such a way as to oppose the change.

7. Phase-inverter Circuits. Sweep voltages are usually applied to the deflection plates of electrostatic cathode-ray tubes in push-pull because of the defocusing effects that result when the sweep voltage is applied to only one plate of a pair. To obtain push-pull sweep voltages, a *phase-inverter* circuit is used. Four types of phase-inverter circuits common in radar indicators are as follows:

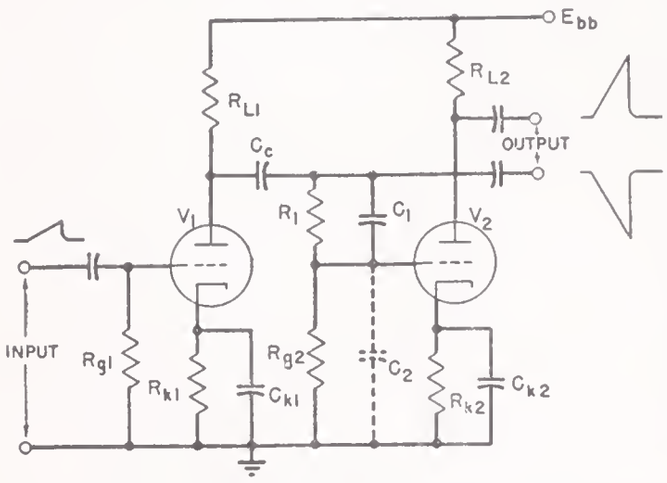


Fig. 11. Attenuator-and-amplifier phase inverter.

Attenuator and Amplifier. The circuit of Fig. 11 uses two identical amplifiers coupled by a voltage divider consisting of R_1 and R_{g2}. A small rise in voltage at the grid of V_1 causes a large drop in plate voltage which is applied to one of the deflection plates. Voltage changes at the plate of V_1 are also coupled through the voltage divider to the grid of V_2. The values of R_1 and R_{g2} are chosen so that the attenuation of the voltage divider equals the amplification of V_2. Plate-voltage changes of V_2 then have the same magnitude as those of V_1, but are of opposite polarity.

Often a small capacitor is placed in parallel with R_1 to make the voltage-divider ratio independent of the frequency of the Fourier components (see Art. 2, Chap. VI) of sweep-voltage waveform. Resistances R_1 and R_{g2} determine the division ratio for the low-frequency components, while the capacitance voltage divider, consisting of C_1 and the input capacitance of V_2, controls the division ratio for the high-frequency components. If the two voltage-division ratios are made equal, all Fourier components of the voltage are attenuated in the same ratio and no distortion of the input voltage for V_2 is caused by the voltage divider.

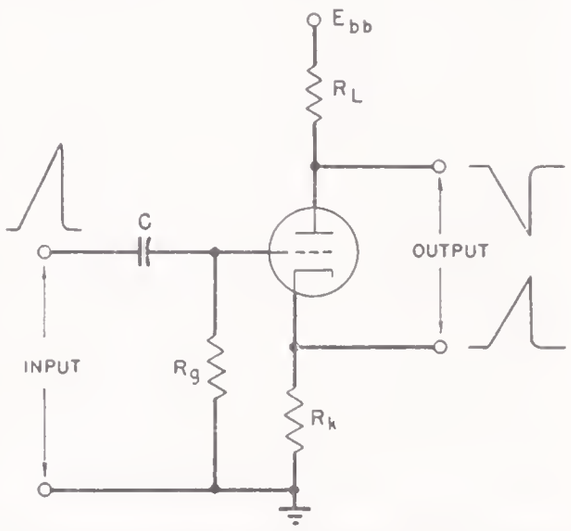


Fig. 12. Single-tube phase inverter.

This form of phase inverter has the disadvantage that the sweep-voltage waveform is distorted not only by nonlinearities of tube V_1 but also by nonlinearities of the inverter tube V_2.

Single-tube Phase Inverter. A phase inverter that uses both plate and cathode load resistors is drawn in Fig. 12. In this circuit a rise in voltage applied to the grid causes a rise in voltage at the cathode and a drop in voltage at the plate. Resistors R_L and R_k are made equal; therefore, the

plate current of the tube produces equal-magnitude voltages across the resistors, and the two output waveforms are of equal amplitude. The amplification of the circuit, from input to either output terminal, is less than unity because, as in a cathode follower, the cathode-voltage change must be less than the change of grid-to-ground voltage. This phase inverter has the disadvantage that the maximum amplitude of each output signal can be only half as great as the maximum amplitude of the output from an ordinary amplifier operated from the same plate supply voltage.

Self-balancing Phase Inverter. The phase-inverter circuit of Fig. 13 is used with the sweep-generating circuits of Figs. 9 and 10. Tube V_1 is the cathode follower of the sweep-generating circuit, and tube V_2 is an

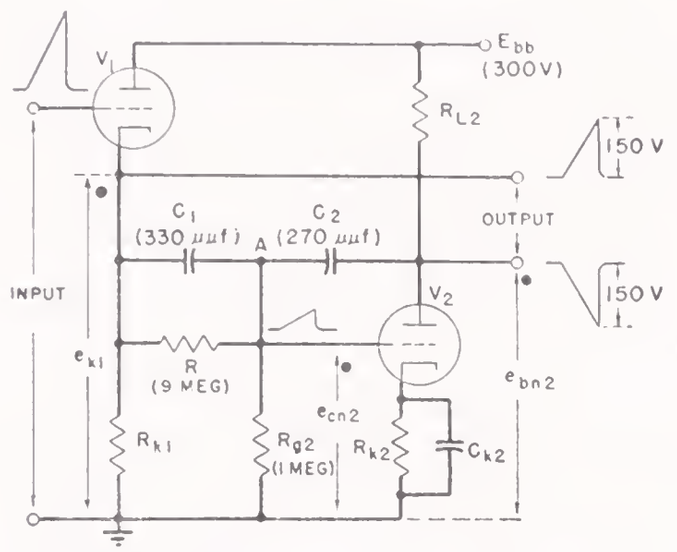


FIG. 13. Self-balancing phase inverter.

inverting amplifier like tube V_2 in Fig. 11. A special balancing circuit is used instead of an attenuator to supply the input voltage to tube V_2 with the result that distortion of the sweep voltage caused by the inverting tube is greatly reduced.

Voltage e_{k1} at the cathode of V_1 is a positive-going saw-tooth wave and is applied to one of the cathode-ray-tube deflection plates. The voltage e_{bn2} at the plate of V_2 is applied to the other deflection plate. In order that a push-pull sweep voltage may be obtained, the wave of e_{bn2} must be a negative-going saw tooth equal in amplitude to the wave of e_{k1}. The grid-voltage wave e_{cn2} should therefore be a positive-going saw tooth; and if, for example, the amplification of V_2 is 10, the amplitude of the grid wave of V_2 should be one-tenth that of the plate wave of V_2 and cathode wave of V_1.

One method of obtaining the desired grid voltage would be to open-circuit capacitors C_1 and C_2 and use resistors R and R_{g2} as a 10:1 attenuator, exactly as in the circuit of Fig. 11. Although such a circuit is sometimes used, this type of connection causes excessive attenuation of the

high-frequency components of the saw-tooth wave (because of the input capacitance of V_2), and nonlinear amplification by tube V_2 causes additional distortion of the sweep waveform.

Another method of obtaining the desired wave of e_{cn2} would be to open-circuit resistor R and use capacitors C_1 and C_2 as a voltage divider between e_{k1} and e_{bn2}. If e_{k1} rises by 150 volts during the sweep, e_{bn2} should fall by 150 volts; and the capacitor sizes must be so chosen that the voltage e_{cn2} at point A rises by $^{150}/_{10} = 15$ volts if the gain of V_2 is 10. The voltage across capacitor C_1 must therefore change by $150 - 15 = 135$ volts during the sweep, and the voltage across C_2 by $150 + 15 = 165$ volts. If R_{g2} is so large that its current causes negligible change of capacitor charge during the sweep, then the capacitance values must be inversely proportional to the desired voltage changes, or $C_1/C_2 = {}^{165}/_{135}$. Thus the capacitances might be $C_1 = 330$ $\mu\mu$f and $C_2 = 270$ $\mu\mu$f, as indicated in the figure. In general terms, if A is the amplification of tube V_2 and ΔE the change of e_{k1} or e_{bn2} during the sweep, the required change of voltage across C_1 during the sweep is $\Delta E(1 - 1/A)$, and that across C_2 is $\Delta E(1 + 1/A)$; and therefore the capacitance ratio should be

$$\frac{C_1}{C_2} = \frac{\Delta E(1 + 1/A)}{\Delta E(1 - 1/A)} = \frac{A + 1}{A - 1} \tag{1}$$

The reason for connecting C_1 and C_2 as in Fig. 13, instead of forming a 10:1 attenuator for e_{k1} as in Fig. 11, is that the cathode-to-plate connection provides a self-balancing action that tends to maintain e_{k1} and e_{bn2} equal in amplitude despite changes in amplification of V_2. If the output of V_2 decreases, the amplitude of the voltage wave at point A *increases*, thus increasing the input to V_2 and opposing the decrease of the output. Consider, for example, the loss of amplification that is necessary to reduce the amplitude of the e_{bn2} wave from 150 to 140 volts. A decrease of e_{bn2} of 140 volts during the sweep and a rise of e_{k1} of 150 volts cause the voltage across C_1 to change by

$$\frac{270}{270 + 330}(150 + 140) = 130.5 \text{ volts}$$

The corresponding rise of e_{cn2} is $150 - 130.5 = 19.5$ volts, and the amplification of V_2 is $140/19.5 = 7.18$. Thus a loss of amplification of $(10 - 7.18)/10 = 0.282$, or 28.2 per cent, causes the output voltage to decrease by only $(150 - 140)/150 = 0.0667$, or 6.67 per cent. The self-balancing effect produced by the capacitors not only equalizes the over-all amplitudes of the sweep waveforms but also tends to correct the waveform of e_{bn2} at each point along the saw tooth. In this way, distortion of the e_{bn2} waveform caused by nonlinear amplification of tube V_2 is greatly reduced.

Capacitors C_1 and C_2 must be fairly small, because they affect the cathode-follower operation in the same way as would a capacitance shunted across R_{k1}; that is, they attenuate the high-frequency components of the sweep waveform. At the same time, C_1 and C_2 operate with R_{g2} as an R-C coupling circuit; and because of the small size of the capacitors, low-frequency components of the wave supplied to the grid of V_2 would be attenuated if R were open-circuited. In terms of the actual sweep waveform, current in R_{g2} would cause some change of the capacitor voltage, with the result that e_{cn2} would rise along an exponential curve during the sweep, falling more and more below the desired straight line as the sweep progresses. Furthermore, with R open-circuited, the average

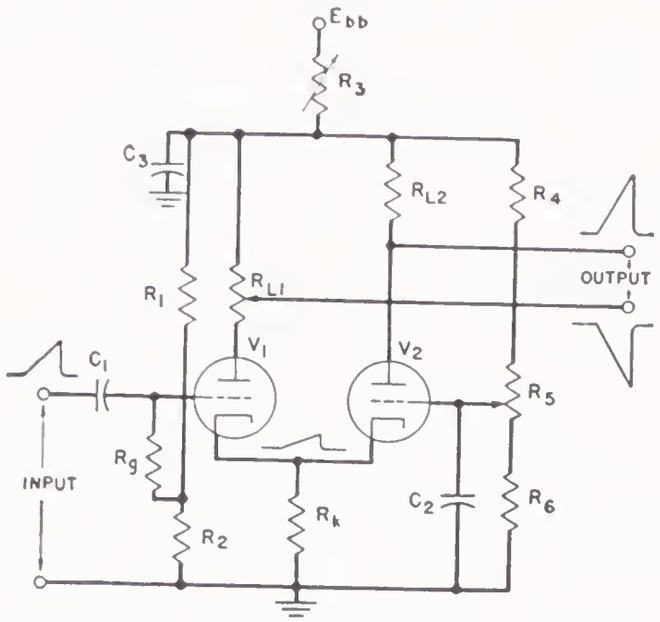


FIG. 14. Cathode-coupled phase inverter.

voltage across R_{g2} would be zero; and, therefore, the value of e_{cn2} between sweeps would change when the sweep duration changed. To prevent these effects, the actual circuit employs both the self-balancing capacitance divider (C_1 and C_2) to supply the high-frequency components of e_{cn2} and the resistance divider (R and R_{g2}) to supply the low-frequency and d-c components of e_{cn2}. As in the compensated voltage divider of Fig. 11, the ratios of the two dividers (measured from e_{k1} to e_{cn2}) must be equal in order that all the Fourier components of the waveforms may be attenuated by the same amount.

Instead of capacitors C_1 and C_2, a resistance voltage divider is sometimes connected between the cathode of V_1 and the plate of V_2, and the connection from point A to the grid of V_2 is made through a coupling capacitor. The self-balancing action of such a circuit is essentially the same as that obtained with a capacitance voltage divider.

Cathode-coupled Phase Inverter. In the phase-inverter circuit of Fig. 14, a rise of voltage at the grid of V_1 produces a small rise in the cathode

voltage and a large drop in the plate voltage. A negative-going sweep voltage is therefore obtained at the plate of V_1. The change in cathode voltage is the effective input signal of V_2. This signal makes the grid-to-cathode voltage negative and causes a large rise in voltage at the plate of V_2. The positive-going half of the push-pull sweep is thus obtained at the plate of V_2.

The gain from the input to either output terminal is approximately one-half the gain that V_1 would have if its cathode were connected to a fixed bias voltage. Consider first that $R_k = 0$ and that V_1 is operated with a fixed negative bias. The amplification of V_1 has a value determined by the magnitude of the plate load resistor, and the amplification of V_2 is zero. If the value of R_k is now increased until the grid-to-cathode signal voltage of V_1 is one-half its original value, the gain of this stage will be reduced by a factor of approximately one-half. However, the grid-to-cathode signal voltage of V_2 is now equal to that of V_1, with the result that two output voltages are obtained, each of which has an amplitude of approximately one-half the output of a single tube operating with a fixed bias. Since changes in plate current are in opposite directions in the two tubes, any change in cathode voltage requires that the two plate-current changes be unequal in magnitude. Equal output voltages at the plates of the tubes can therefore be obtained only by making R_{L2} larger than R_{L1}, or by using only part of the output available from the plate of V_1.

Where the cathode-coupled type of phase-inverter circuit is used, the deflection plates of the cathode-ray tube are often connected directly to the plates of V_1 and V_2. Thus the between-sweeps plate voltages of V_1 and V_2 determine the starting point of the sweep on the screen. The between-sweeps voltages may be controlled by connecting the grid of V_2 to a source of variable direct voltage, as shown in the figure. Varying the grid-to-ground voltage of V_2 has opposite effects on the between-sweeps voltages at the plates of V_1 and V_2 for exactly the same reason that the saw tooth applied to the grid of V_1 produces opposite-polarity waves at the plates of V_1 and V_2. Potentiometer R_5 can therefore be used to control the position of the sweep on the indicator screen. Because of resistor R_k, a signal applied to the grid of *either* tube causes an output voltage at the plates of *both* tubes. In Fig. 14 the alternating components of the desired output wave are applied to the grid of V_1, and the direct component to the grid of V_2. Both alternating and direct components appear in each plate-voltage waveform.

Rheostat R_3 controls the effective plate supply voltage for the circuit and permits the average voltage of the deflection plates with respect to the second anode to be varied. Rheostat R_3 can therefore be used as an astigmatism control (see Art. 1, p. 219). The positioning of the pattern

is not affected by this adjustment because the direct-voltage level of both waveforms is changed by the same amount and in the same direction.

Some radar sets employ the cathode-coupled phase inverter to supply video signals together with range markers, a range step, or other auxiliary voltages to the vertical-deflection plates. When this type of presentation is desired, the echo signals are applied to the grid of one of the phase-inverter tubes and the auxiliary voltages are applied to the grid of the other tube. Because of the common cathode resistor, a mixture of all signals appears at each amplifier-tube anode.

8. Delayed Sweeps and Expanded Sweeps. When long-range sweeps are used, accurate measurement of range becomes impossible because very small increments of range cannot be observed. The accuracy with which small range differences can be observed is increased if a small section of

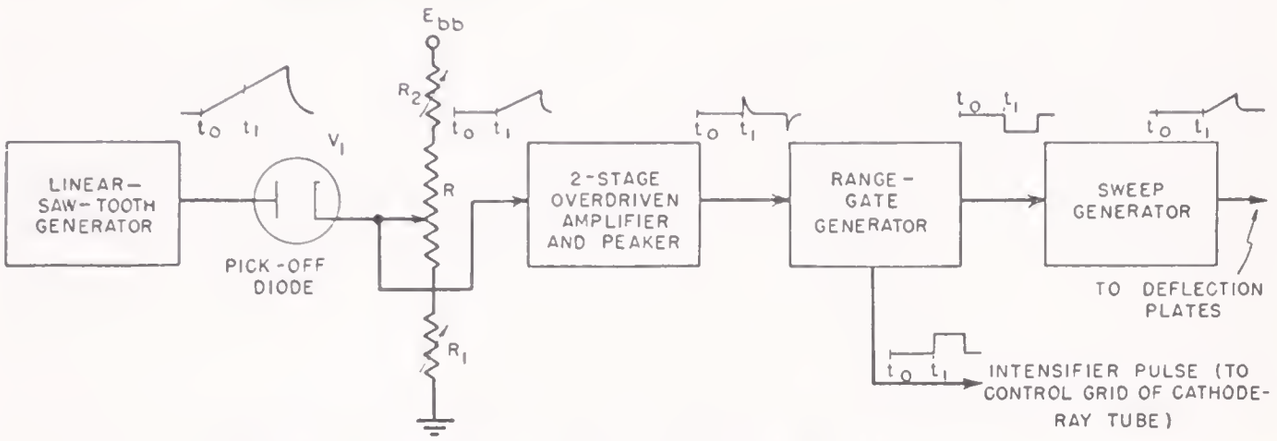


FIG. 15. Typical circuit for providing a delayed sweep.

the sweep is expanded to cover the full length of the screen (see Chap. I, p. 29). Figure 15 illustrates the essential elements of a circuit that expands the latter portion of a range sweep.

The linear saw-tooth generator may be any of the circuits described; however, the bootstrap circuit is most commonly employed. The cathode of the pick-off diode V_1 is maintained at a positive direct voltage determined by the position of the tap on R. At the time of the transmitted pulse, the plate voltage of V_1 begins to rise linearly. However, tube V_1 does not conduct until sufficient time has elapsed for its plate voltage to rise above its cathode voltage (compare the diode step-generator circuit of Art. 12, Chap. II). As shown in the diagram, V_1 conducts at time t_1 and impresses a rising voltage upon the input of a two-stage overdriven amplifier and peaker circuit. The output of the peaker is a sharp positive pulse at time t_1. This pulse is used to trigger the range-gate generator which supplies to the sweep generator a negative gate pulse of duration equal to the desired expanded-sweep duration. The sweep generator, therefore, provides a sweep delayed by a time $t_1 - t_0$ relative to the transmitted pulse.

The voltage-divider device R may be either a potentiometer or a stepped voltage divider. When a potentiometer is employed, it is sometimes helically wound to give a relatively small change in resistance for a large angle of rotation of the shaft. The delay time in yards is usually read from a calibrated dial attached to the shaft. In order to make the dial read correctly, two adjustments are required. With the range-delay dial set on zero, the cathode voltage of V_1 is adjusted by means of R_1 (called the zero control) so that the indicator sweep begins at the time of the transmitted pulse. Then the range-delay dial is set on the maximum-delay point, and the starting time of the indicator sweep is adjusted by means of R_2 (called the slope control) to give the delay indicated on the dial. Range markers are commonly used to measure the delay time. Because the two calibrations are not independent, successive adjustments of R_1 and R_2 must be made to obtain the correct settings.

Where a stepped voltage divider is used, taps are provided to indicate definite delay intervals—such as 10,000 yd, 20,000 yd, and so on.

In a variation of the circuit of Fig. 15, the linear-saw-tooth generator, pick-off diode, and calibrated voltage divider are replaced by either a cathode-coupled multivibrator or a phantastron delay circuit. The delay may be varied when using the cathode-coupled multivibrator by changing the direct grid voltage of the normally off tube (Art. 15, Chap. II) and in the phantastron by use of a control diode (Art. 16, Chap. II).

9. Indicators Used with Antenna Switching. Antenna switching is ordinarily employed only in uhf radar equipment where antenna beams are broad and a means of improving angular accuracy is essential. The narrow beams obtained in microwave sets provide sufficient accuracy for most purposes if the antenna position for maximum echo strength is used to indicate direction. For high-accuracy applications, such as fire control, microwave sets are used either with antennas that produce exceedingly sharp beams or with conical scan (generally used in automatic-tracking sets).

One example of antenna switching used in a uhf radar system is indicated by Fig. 16. The two Yagi antennas (see the discussion of these in Art. 12, Chap. XIII) are oriented so that the axes of their beams are in the horizontal plane but separated in azimuth by an angle about equal to the beam angle. The dependence of the intensity of signals from each antenna on azimuth angle is indicated by the radiation patterns[1] in Fig. 16. Switching is necessary because the beams from these antennas are rather wide—the beam angle may be 15 to 50 deg.

A motor-driven switch alternately connects each of the two antennas to the receiver input. The video signals from the receiver output are switched at the same time, so that signals received by one antenna are

[1] For a discussion of radiation patterns see Art. 5, Chap. XIII.

applied to one deflection plate and signals from the other antenna are applied to the opposite plate. When a range-sweep voltage is impressed upon the other set of plates, echo signals cause deflections to both sides of the range trace, and a Type L presentation is produced. The antenna assembly can be rotated in azimuth and is "on target" when the two signal deflections have equal magnitudes. The switching rate is usually considerably less than the pulse repetition frequency, so that several echo pulses are obtained in each position of the switch, yet it is high enough to prevent noticeable flicker of the image on the indicator screen.

A separate transmitting antenna having a broad beam directed midway between the beam axes of the two receiving antennas may be used in the system of Fig. 16. The azimuth accuracy is then determined by the

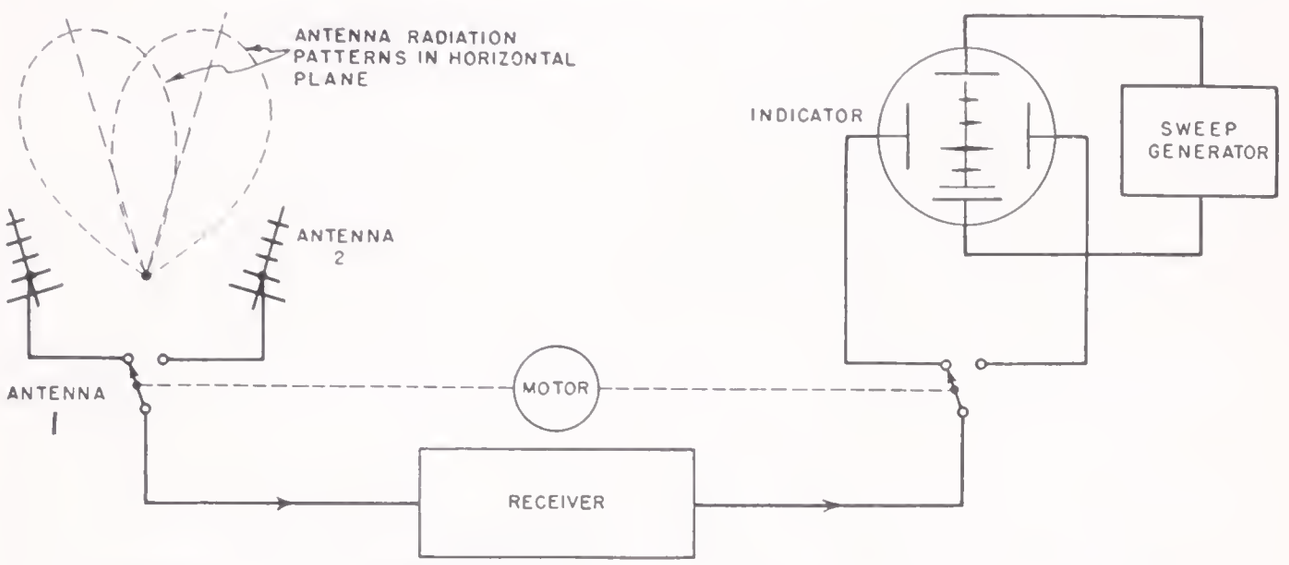


FIG. 16. Block diagram of circuit connections for Type L presentation.

sharpness of the receiving-antenna patterns. Another possibility is to employ a t-r device (Arts. 10 to 14, Chap. XII) so that the antennas of Fig. 16 may be used for both transmission and reception. Transmitter and receiver are switched simultaneously from one antenna to the other, and the directional properties of the antennas are used in both transmission and reception.

Several of the functions of the antenna-switching system of Fig. 16 may be carried out in different ways. Three possibilities are illustrated in Fig. 17. (1) Instead of two antennas having different orientations, a single antenna is used, and the position of the main lobe of the radiation pattern is shifted by changing the connections to the antenna. (2) Instead of a motor-driven mechanical switch, an electronic switching method is employed; the square-wave generator causes the two r-f amplifiers to operate alternately. (3) Instead of the Type L presentation of signals from the two lobes in a back-to-back display, the Type K presentation is used, and the two sets of echo signals appear side by side.

The receiving antenna of Fig. 17 is a uhf broadside array (Fig. 6 of

Chap. I, p. 9) having two output connections. Radiation pattern *A* applies if echo signals are received from one output connection through r-f amplifier *A*, and pattern *B* applies if the other output connection and amplifier *B* are used. A separate transmitting antenna is employed. Gate pulses from the square-wave generator control the r-f amplifiers so that the receiver obtains pulses from the two amplifiers in turn. The square-wave frequency is low, and several repetition cycles of the radar set occur while each amplifier is operating.

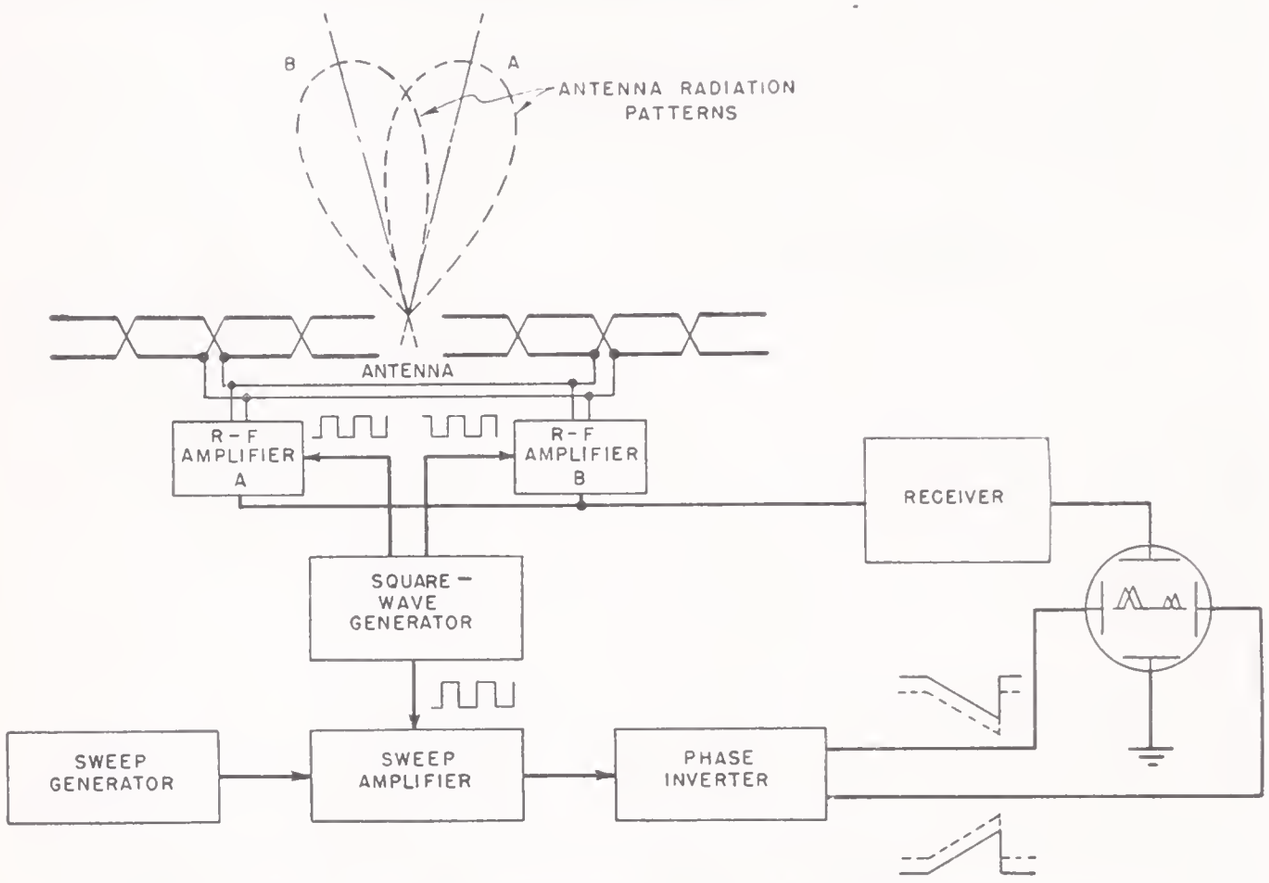


Fig. 17. Block diagram of circuits used in obtaining Type K presentation.

The square wave is applied also to the sweep-amplifier circuit to raise and lower slightly the direct-voltage level of the saw-tooth waveform. The effect is as though the sweep position control were turned back and forth at the switching rate. The position of the sweep trace changes each time the signal path shifts from one r-f amplifier to the other, and the echo pulses from the two antenna patterns appear side by side. An adjustment, the *spread control*, determines the amplitude of the square wave applied to the sweep amplifier and thus controls the separation of the signals. To determine the direction of an object, the operator rotates the entire antenna assembly until the two echo pulses obtained from the object are of the same amplitude.

10. Type J Presentation. A circular sweep trace with radial deflections indicating echo signals, as in Fig. 6, is required for Type J presentation. The transmitted pulse causes a deflection at the start of the sweep, usually the top point of the circle, and range is measured by the distance

along the circle, ordinarily taken clockwise, from the transmitted-pulse deflection to the echo indication. The advantages of Type J over Type A presentation are (1) the greater length of the range trace for a fixed indicator-tube diameter and (2) the ease with which an accurately linear

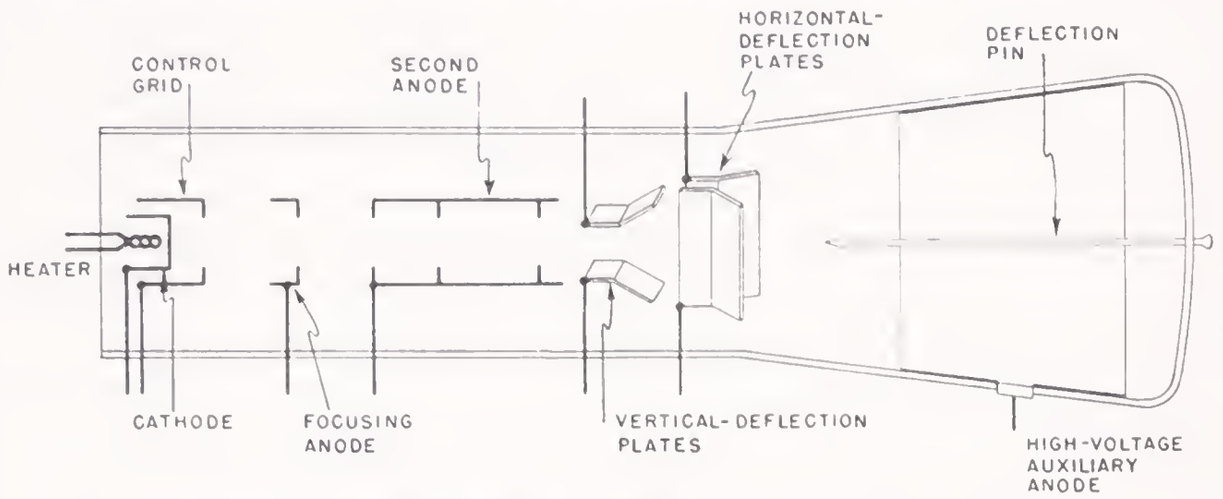


FIG. 18. Cathode-ray tube used for Type J indication.

range scale may be obtained—for constant speed of the spot, sine-wave rather than saw-tooth sweep voltages are required. For high-accuracy range determination, two Type J indicators may be used, one being an expanded presentation of a short section from the full range interval of the other.

Type J presentation makes use of an electrostatic cathode-ray tube of special design. A pin-shaped electrode penetrates the center of the screen, as shown in Fig. 18. Sweep voltages are applied to the horizontal- and vertical-deflection plates and cause the circular motion of the spot. Negative voltages applied to the center pin (with the second anode at ground potential) cause an electric field that deflects the electron beam away from the pin, thus producing echo indications.

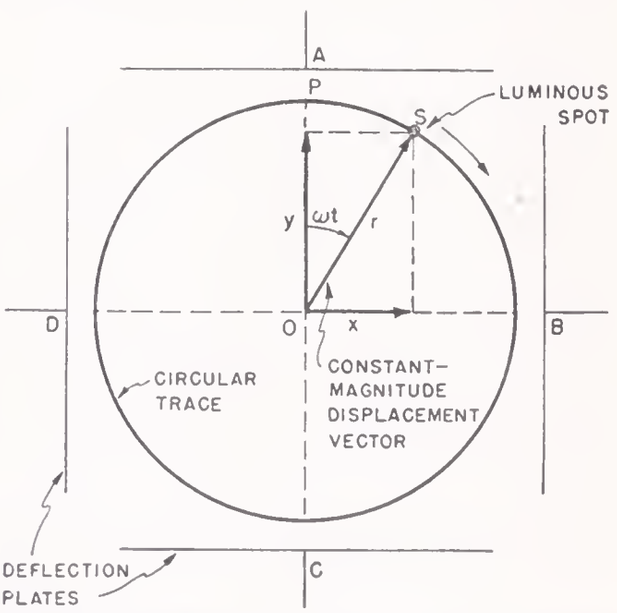


FIG. 19. Resolution of circular sweep motion into horizontal and vertical components of deflection.

The sweep voltages required to produce the circular trace may be determined from Fig. 19. Vector OS indicates the displacement of the luminous spot from the center of the screen. Because the spot follows the circular trace at constant velocity, the length r of OS remains constant and the angle POS changes at a constant rate. If the spot starts from point P at time $t = 0$ and moves clockwise with constant angular velocity ω, then the angle POS is ωt.

The vertical component of the displacement vector OS, which is produced by the voltage applied to the vertical-deflection plates, is $y = r \cos \omega t$. Consequently, a sinusoidal vertical-deflection voltage is required. Similarly, a sine wave of voltage must be applied to the horizontal-deflection plates to produce the deflection $x = r \sin \omega t$. The horizontal and vertical deflections have equal maximum values and are 90 deg apart in phase. The two deflecting voltages must also be 90 deg apart in phase, but of slightly different amplitudes to allow for slightly different deflection sensitivities in the horizontal and vertical directions. The production of a rotating vector, such as the displacement OS, by adding two sinusoidally varying vectors of constant direction is of importance in a number of radar devices.[1]

The direction of rotation of OS depends upon whether the horizontal-deflection voltage leads or lags the vertical-deflection voltage. The deflection voltages are usually applied to the plates in push-pull to avoid defocusing effects; that is, when a positive voltage is applied to plate A of Fig. 19 (relative to second-anode potential), an equal-magnitude negative voltage is applied to plate C. The four deflection-plate-to-anode voltages therefore form a four-phase system, with 90-deg phase differences between the voltages at A and B, at B and C, and so on. For clockwise spot rotation in Fig. 19, the phase sequence is $ABCD$ (voltage at B lags that at A, at C lags that at B, and so on), corresponding to the 90-deg lag of the deflection $x = r \sin \omega t$ relative to the deflection $y = r \cos \omega t$. Thus the spot passes the deflection plates *in the order of the phase sequence*, $ABCD$.

The time for the spot to traverse the circle is the time of one cycle of the sine-wave deflection voltages. Thus the frequency required for these voltages depends upon the maximum range of the sweep to be used. The sweep voltage must be synchronized with the repetition cycle of the transmitter as in any other form of presentation. Synchronization may be accomplished by either of two methods: (1) An intermittent sine wave from which the deflection voltages are derived can be generated by a ringing circuit. Oscillations in the ringing circuit are started each time the transmitter is pulsed and are made to decay rapidly at the end of the useful range interval. (2) A continuously operating sine-wave oscillator can be used as master oscillator to time the transmission of pulses and also as a sweep oscillator to generate the deflection voltages. The second method uses simpler sweep circuits but requires that the repetition period be exactly equal to or an exact multiple of the range interval.

11. Ringing Circuit for Type J Presentation. A useful circuit with the first timing method is illustrated by Fig. 20. Tube V_1 and the L-C circuit function as a ringing circuit (see Art. 19, Chap. II); an oscillating

[1] For examples, see Art. 22 of this chapter and Art. 18 of Chap. V.

voltage is obtained across the L-C circuit during the time V_1 is cutoff by a negative input pulse. Tube V_2 and the L-C circuit form a Hartley oscillator. The feedback adjustment is made so that oscillations are maintained at constant amplitude while V_1 is cutoff but damped rapidly when V_1 conducts. The ringing-circuit voltage is applied through a push-pull amplifier to the horizontal-deflection plates. The same voltage

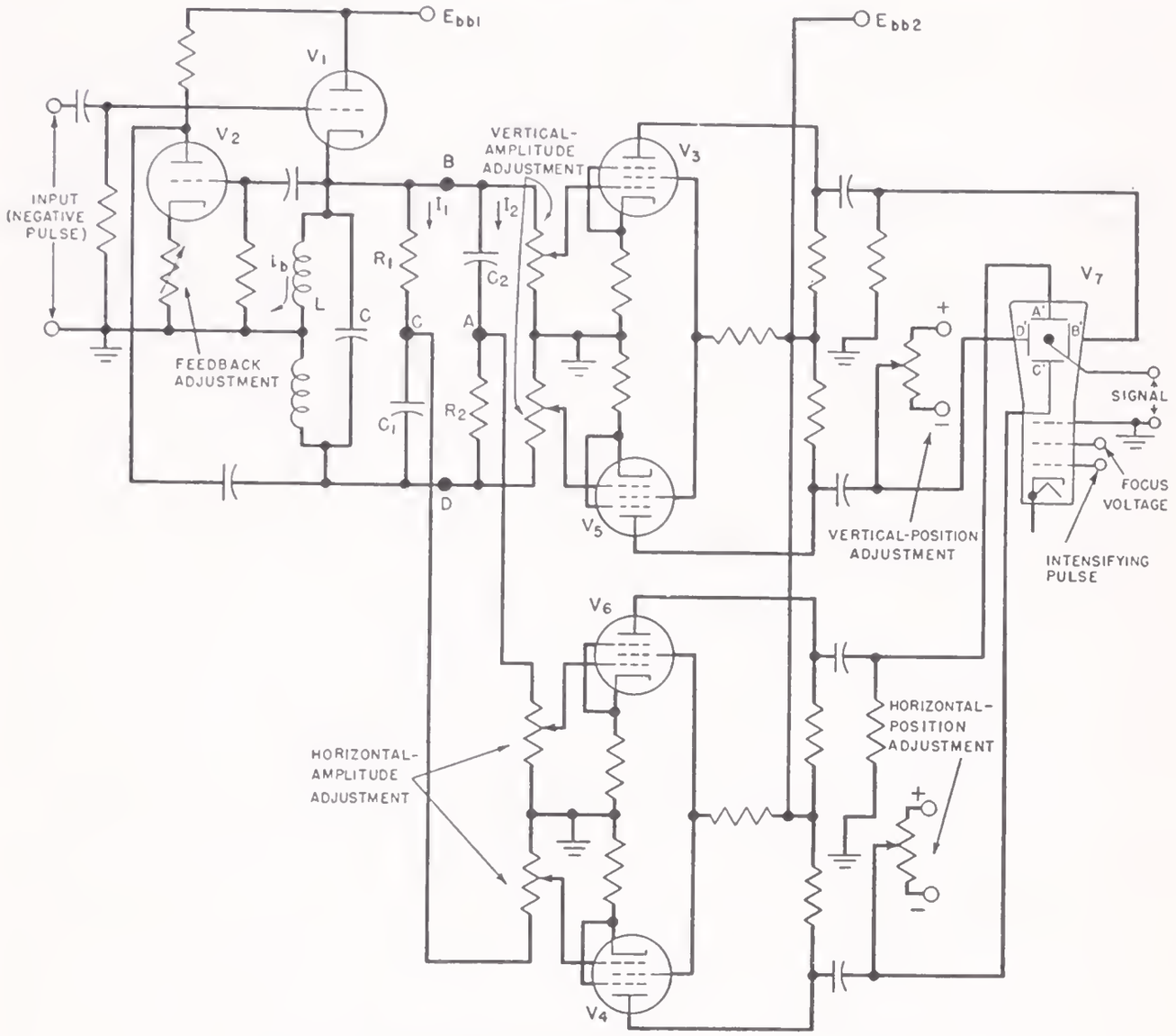


FIG. 20. Range-sweep circuit for Type J indicator.

is shifted 90 deg in phase by an R-C network, then amplified and applied to the vertical-deflection plates.

A difference between the ringing circuit of Fig. 20 and the basic one in Art. 19, Chap. II, is that current initially flows through only half of the inductor in Fig. 20, as indicated by the i_b arrow. The effect of cutting off tube current can be explained simply if the two halves of L, which are wound on the same magnetic core, are assumed to be perfectly coupled. As in the transformer with no leakage inductance (Art. 20, Chap. II), the current in half of L *can* change instantly, *provided* that current in the other half of L makes a simultaneous change so that the total ampere-turns remain unaltered. Currents in the two parts of L must become equal

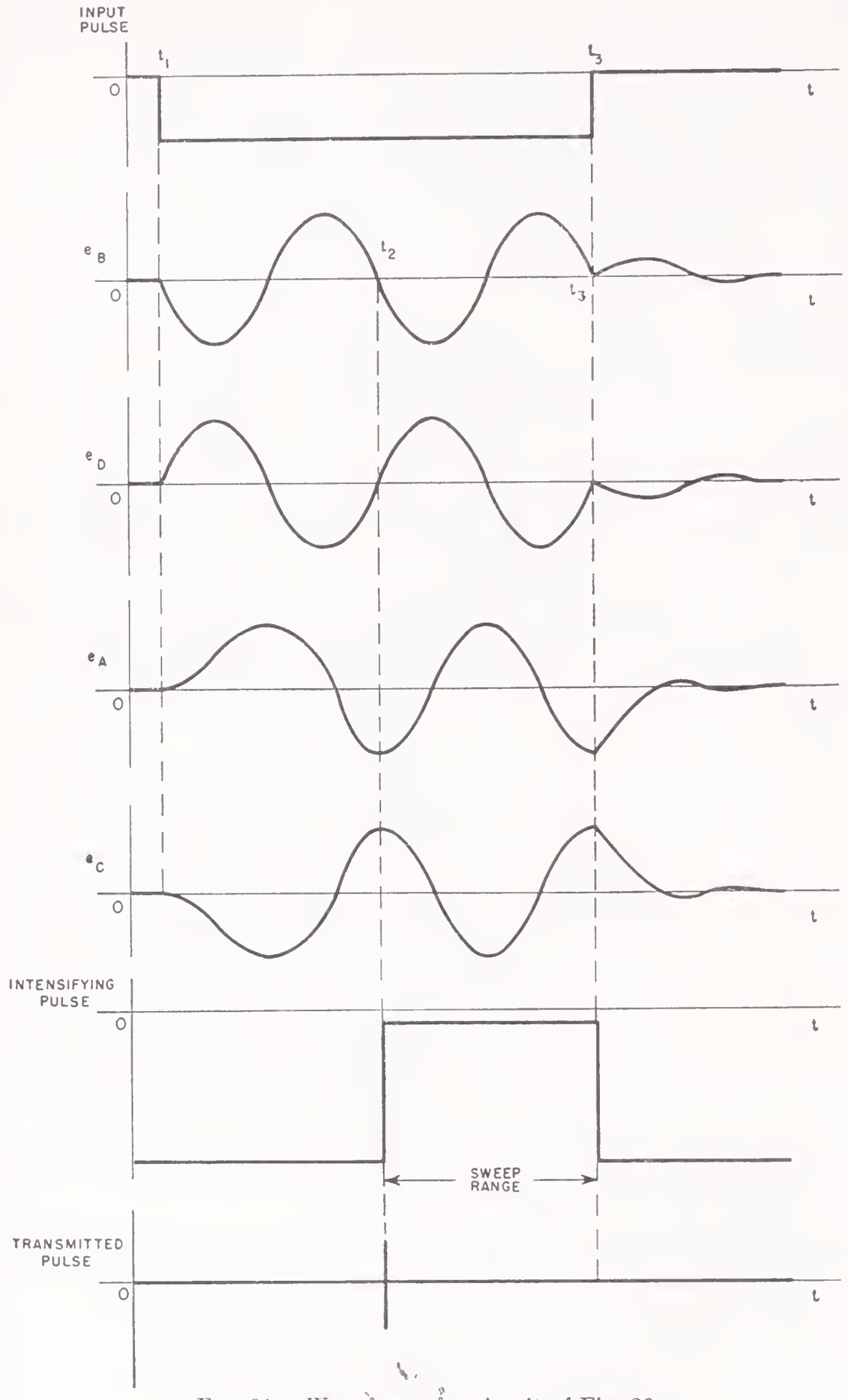


FIG. 21. Waveforms for circuit of Fig. 20.

when the tube is cutoff. Because the tap point is at the center of L, the initial current of i_b in one half of L is changed instantly to $\frac{1}{2}i_b$ in all of L, in order to maintain the total ampere-turns constant. From the current $\frac{1}{2}i_b$, the oscillatory voltage across the L-C circuit can be determined as was done in Art. 19, Chap. II. Because the center point of L is connected to ground, a push-pull output voltage is available. The voltage at point B (with respect to ground) is one-half the resonant-circuit voltage and begins with a decrease from zero to a negative peak value, as in the simple ringing circuit. The voltage at point D has the same magnitude as that at B but is of opposite polarity.

A difficulty of this circuit is that the phase-shifting network operates correctly only in the steady state. When oscillations first begin, exponential transients in the R-C network modify the sinusoidal voltages at points A and C of Fig. 20. These transients decay to negligible proportions during the first cycle of operation. To avoid the effect of these transients, the ringing circuit is allowed to generate one cycle of its oscillation before the transmitter is pulsed. The spot would trace a distorted circle in this interval, but the indicator is blanked so that no trace appears. Time relations are indicated by the waveforms of Fig. 21. At time t_2, the start of the second cycle of oscillation of the ringing circuit, the transmitter is pulsed and an intensifying pulse is applied to the grid of the cathode-ray tube. The intensifying pulse causes the spot to remain visible during the single circular trace corresponding to the second cycle of the ringing oscillation. At time t_3, the cathode-ray tube is blanked and the ringing oscillation ended.

Operation of the phase-shifting network during the steady-state conditions of the second cycle may be explained if complex numbers are used to represent phases and amplitudes of the sinusoidal currents and voltages in the circuit.[1] The currents through the two branches of the phase shifter are represented by the complex numbers I_1 and I_2 (see Fig. 20). The voltages with respect to ground at the points A, B, C, and D in Fig. 20 are represented by the complex numbers E_A, E_B, E_C, and E_D, respectively. Instantaneous values of these same voltages are shown as e_A, e_B, e_C, and e_D in Fig. 21.

In Fig. 22, these complex numbers are plotted in the complex plane.[2] The phase of voltage e_B is taken as the zero-phase reference, so that E_B is plotted on the positive real axis and E_D along the negative real axis. The

[1] The use of complex numbers in sine-wave calculations is explained in most texts on a-c circuits. See, for example, E. E. Staff, M.I.T., *Electric Circuits*, 256–336 (John Wiley & Sons, Inc., New York, 1940).

[2] Diagrams of this kind are usually called *vector diagrams*. This name is avoided in this book in order that the word *vector* may be reserved for quantities such as force, velocity, and electric-field intensity, which have a direction of action as well as a magnitude.

total ringing-circuit voltage, which causes the currents I_1 and I_2, is the voltage $E_B - E_D$ and is in phase with E_B as shown. The currents I_1 and I_2 are equal, and each leads the voltage $E_B - E_D$ by 45 deg because the resistors R_1 and R_2 are equal, and the capacitor reactances X_1 and X_2 are made equal to R_1 and R_2 at the frequency of the oscillations. The resistor voltages I_1R_1 and I_2R_2 are in phase with the currents, and the capacitor voltages I_1X_1 and I_2X_2 lag the currents by 90 deg, or lag $E_B - E_D$ by 45 deg. Because $I_1R_1 + I_1X_1$ must equal $E_B - E_D$, as must $I_2R_2 + I_2X_2$, the resistor- and capacitor-voltage points lie on the vertical line through E_B as shown. Reference to the circuit in Fig. 20 shows that

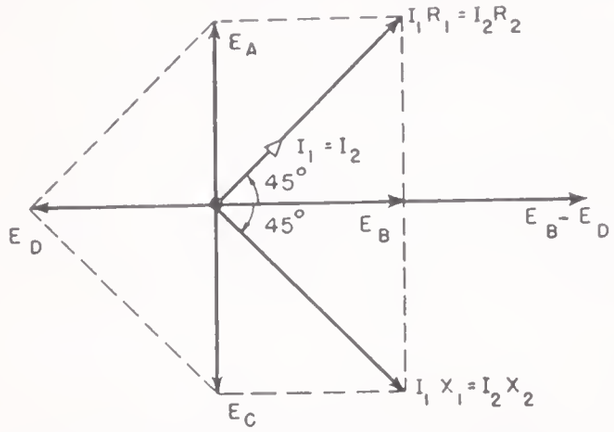


FIG. 22. Complex-number diagram for phase-shifting circuit in Fig. 20.

$$E_A = E_D + I_2R_2$$

and $E_C = E_D + I_1X_1$. Thus the additions indicated by dashed lines in Fig. 22 are used to locate E_A and E_C.

Figure 22 indicates that voltages e_A, e_B, e_C, and e_D constitute the desired set of four-phase voltages. The waveforms of e_A and e_C in Fig. 21 are drawn in accord with the relations obtained from Fig. 22. The phase sequence is $ABCD$, and the sequence of the deflection-plate voltages is $A'B'C'D'$. Thus the spot moves clockwise around the circle. At the start of the visible sweep trace, e_A has its peak negative value, and the voltage at deflection-plate A' in Fig. 21 has its peak positive value. Thus the sweep begins beneath plate A'. The amplitude and position controls permit adjusting the trace to form a circle with its center at the center pin.

12. Crystal-controlled Oscillator for Type J Presentation. When the second synchronizing method is used, the sweep circuit is merely a sine-wave oscillator and a phase-shifting network. Often the oscillator is crystal-controlled to ensure frequency stability and accuracy of range measurement. The phase-shifting network can be the same as in Fig. 20, but other circuits are often used. Figure 23 shows a crystal-controlled oscillator used with a phase-shifting network comprising a three-winding, tuned, air-core transformer.

The oscillator of Fig. 23 is a form of tuned-plate tuned-grid oscillator. Capacitor C_1 is connected to the primary of transformer T_1 to form the resonant plate circuit. The quartz crystal is a mechanical resonant system coupled by the piezoelectric effect[1] to the electric circuit, and this

[1] See, for example, W. G. CADY, "Nature and Use of Piezoelectricity," *Elec. Eng.*, **66** (August, 1947), 758–762, or for a discussion in greater detail, W. P. MASON, "Quartz-Crystal Applications," *B.S.T.J.*, **22** (July, 1943), 178–223.

crystal is the tuned circuit for the grid. Feedback occurs through the grid-to-plate capacitance of the tube. The circuit oscillates only for certain tuning adjustments of C_1, and the oscillation frequency is closely controlled by the crystal.

The quartz crystal controls the oscillator frequency more closely than would a parallel L-C circuit for two reasons: (1) The mechanical resonant

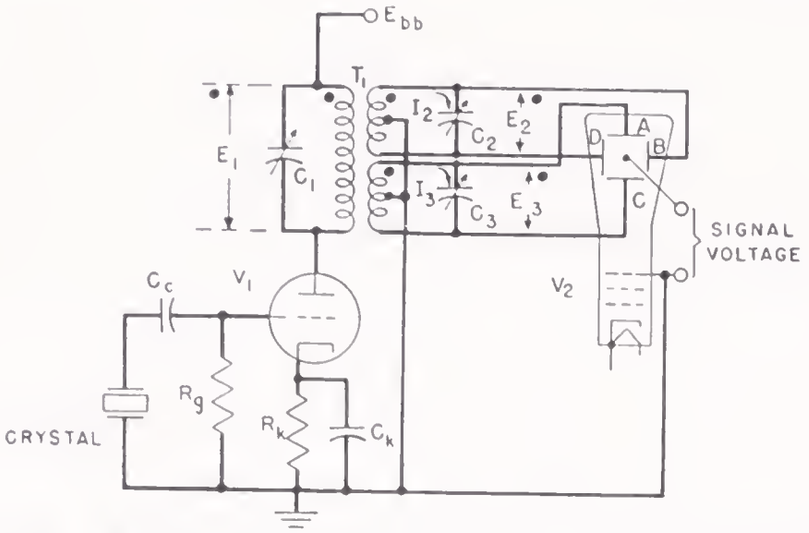


FIG. 23. Alternate range-sweep circuit for Type J indicator.

frequency of a properly cut crystal is very little influenced by the conditions of operation, such as temperature. (2) The frequency of oscillation is always very nearly equal to the frequency of mechanical resonance, because of the nature of the crystal equivalent circuit and because the mechanical resonance is exceedingly sharp—the Q is very much higher than that obtainable with ordinary tuned circuits.

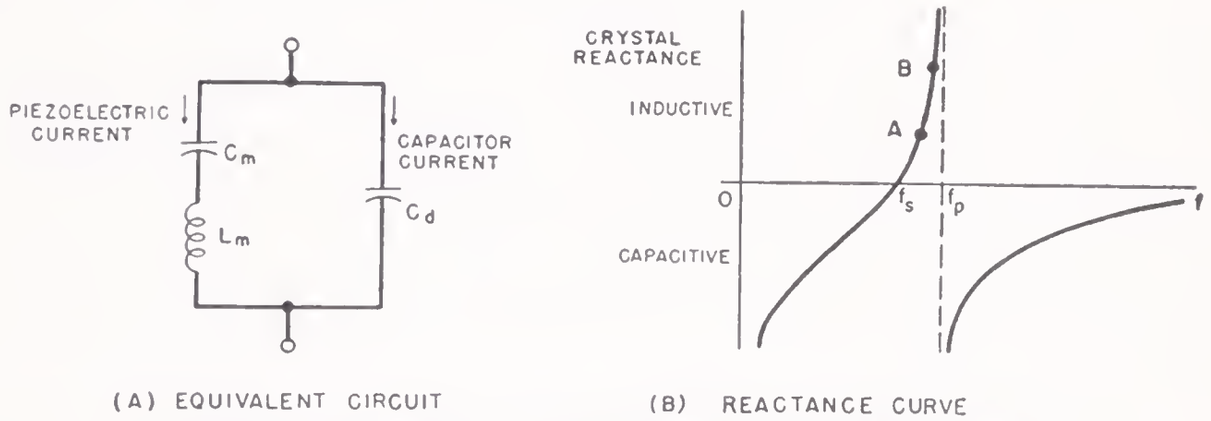


FIG. 24. Equivalent circuit and reactance of quartz crystal.

The equivalent circuit for the crystal, drawn in Fig. 24*A*, accounts for the two components of crystal current. One component, the current in the capacitor formed by the electrodes of the crystal holder with the crystal as dielectric, is the current through C_d in Fig. 24*A*. The other part is the result of charges produced on the crystal surfaces by the piezoelectric effect. This current component depends upon the mechanical resonance of the crystal and is represented in Fig. 24*A* by the current in a circuit tuned to the mechanical resonant frequency.

The plate and grid circuits of a tuned-plate tuned-grid oscillator have inductive reactances at the frequency of oscillation, and the sum of these reactances resonates with the grid-to-plate capacitance.[1] The circuit of Fig. 23 must therefore oscillate at a frequency for which the crystal reactance is inductive. The crystal reactance is plotted as a function of frequency[2] in Fig. 24*B*. At low frequencies, the reactances of both branches of the equivalent circuit are capacitive. At frequency f_s, the series-tuned branch resonates and is effectively a short circuit, thus making the over-all reactance zero. At a somewhat higher frequency f_p, the series-tuned branch has an inductive reactance equal in magnitude to the capacitive reactance of C_d, and parallel resonance is obtained. Between f_s and f_p, the inductive reactance is lower than the capacitive reactance, and the parallel combination of the two reactances is inductive. Above f_p, the reverse is true, and an over-all capacitive reactance is obtained. Only at frequencies within the range between f_s and f_p is the crystal reactance inductive, and only at these frequencies is oscillation possible for the circuit of Fig. 23.

The crystal controls the oscillator frequency accurately for one reason, because capacitor C_d in Fig. 24*A* is very much larger than C_m, so that f_p is only slightly higher than f_s and the range of inductive reactance is small. Additional accuracy of control is obtained because oscillation is possible over only a part of this range, for example, between points *A* and *B* in Fig. 24*B*. If capacitor C_1 in Fig. 23 is adjusted to increase the plate-circuit inductive reactance, the grid-circuit reactance and the frequency of oscillation decrease as the point of operation in Fig. 24*A* moves toward *A*. At points below *A*, the grid-circuit reactance is too small to provide feedback voltage sufficient in magnitude to sustain oscillations. Adjustment of C_1 to decrease the plate-circuit inductive reactance moves the operating point toward *B* and increases the oscillation frequency. Point *B* is located at the frequency for which the grid-circuit inductive reactance resonates with the grid-to-plate capacitance. Higher frequencies of oscillation are impossible. Capacitive plate-circuit reactance would be required, and the feedback voltage would have the wrong polarity to sustain oscillations.

Sinusoidal voltages are produced across the primary and two secondary windings of transformer T_1 in Fig. 23. By adjustment of capacitors C_2 and C_3, the secondary voltages can be made equal in amplitude and 90 deg

[1] The principles of Art. 2, Chap. X, can be applied to the analysis of the tuned-plate tuned-grid oscillator by methods analogous to those employed in Art. 4, Chap. X, for the plate-return oscillator. By this means, the resonator of a tuned-plate tuned-grid oscillator may be shown to comprise the grid-plate capacitance shunted by a series combination of inductive reactances representing the plate and grid circuits.

[2] Reactance curves of this kind are studied in considerable detail in Chap. IX. See especially Arts. 6 and 7.

apart in phase. Phase and amplitude relations may be determined from the equivalent circuit and complex-number diagram of Fig. 25. For simplicity, 1:1:1 turns ratios are assumed, and ideal transformers are omitted from the equivalent circuit. The equivalent circuit is like the pulse-transformer equivalent in Art. 20, Chap. II, except that the circuit of the third winding is added. For air-core transformers, coupling coefficients are low, and therefore the leakage inductances in Fig. 25*A* are important. Transformer losses, neglected in connection with the pulse transformer, are accounted for by resistors R_1, R_2, and R_3 in Fig. 25*A*.

Capacitor C_2 is adjusted to a value somewhat larger than that yielding resonance of the L_{l2}, R_2, C_2 circuit. Therefore, inductive reactance predominates in this circuit, and I_2 lags the magnetizing-inductance voltage E, as shown in Fig. 25*B*. Capacitor C_3 is adjusted to less than the

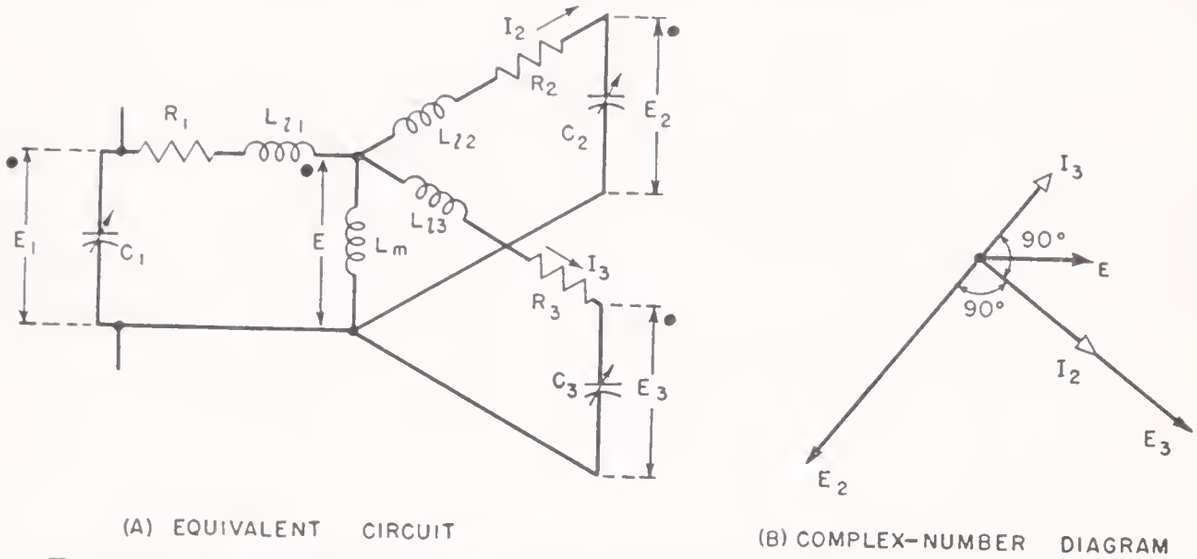


FIG. 25. Amplitude and phase relation in tuned three-winding transformer.

resonant capacitance, and I_3 leads E. The capacitor values are adjusted so that the total phase angle between the two currents is 90 deg. The phase angle between voltages E_2 and E_3 is also 90 deg, because each capacitor voltage lags the corresponding capacitor current by 90 deg. If the two secondary windings are alike, the angle of lag of I_2 is made somewhat less than 45 deg; and the angle of lead of I_3 is correspondingly greater than 45 deg. Thus the circuit of winding 2 is the more nearly resonant, and the magnitude of I_2 is greater than that of I_3. This difference of current magnitudes is necessary for equal magnitudes of E_2 and E_3 because the reactance of C_2 is less than that of C_3.

As indicated in Fig. 23, the center points of the two secondary windings are connected to the second anode of the cathode-ray tube. Thus the required set of four-phase voltages is applied to the deflection plates. For the conditions of Fig. 25*B*, the phase sequence is *ABCD*, and the spot moves clockwise.

Timing circuits actuated by the plate voltage of V_1 are required to control the transmitter so that each pulse is sent out at a time correspond-

ing to a fixed point of the circular sweep. Pulses can be transmitted once each sweep cycle, but more often frequency-dividing multivibrators or blocking oscillators are included in the timing circuits, and several sweep-oscillation cycles occur in the period between transmitted pulses. An intensifying pulse must then be used so that only a single circular sweep trace is visible for each transmitted pulse.

C. SWEEP-CURRENT CIRCUITS

The fields that deflect the electron beam in magnetic cathode-ray tubes are very nearly proportional to the currents flowing in the deflection-coil windings. Therefore, the sweep circuits associated with magnetic tubes must provide *currents*, rather than voltages, which have the proper form

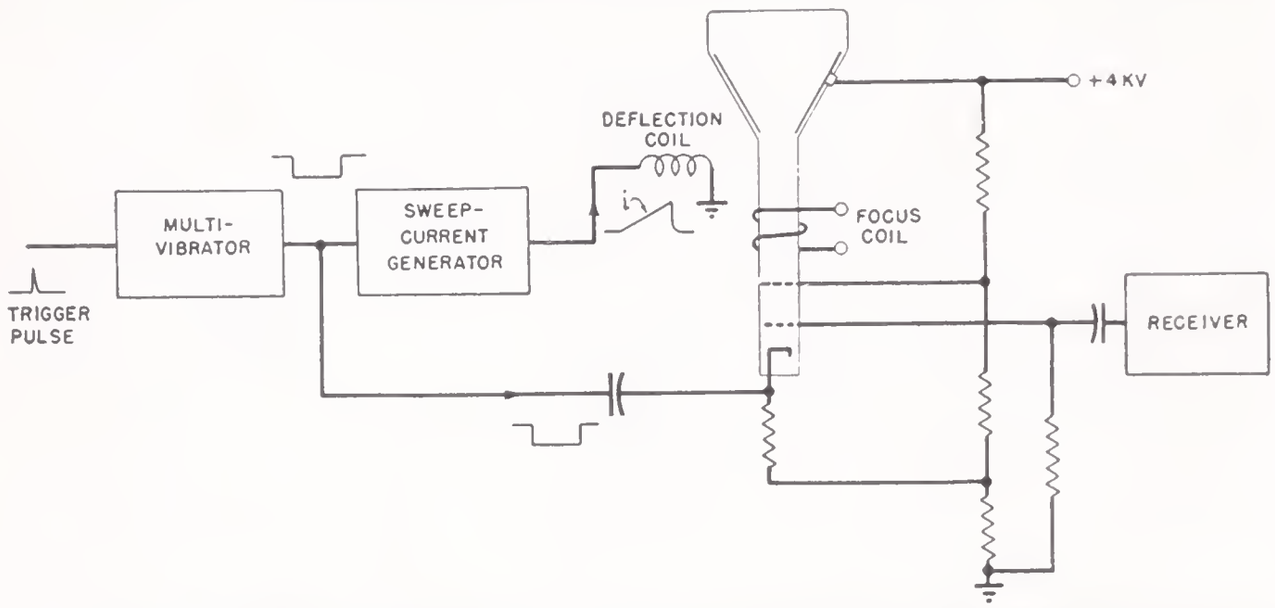


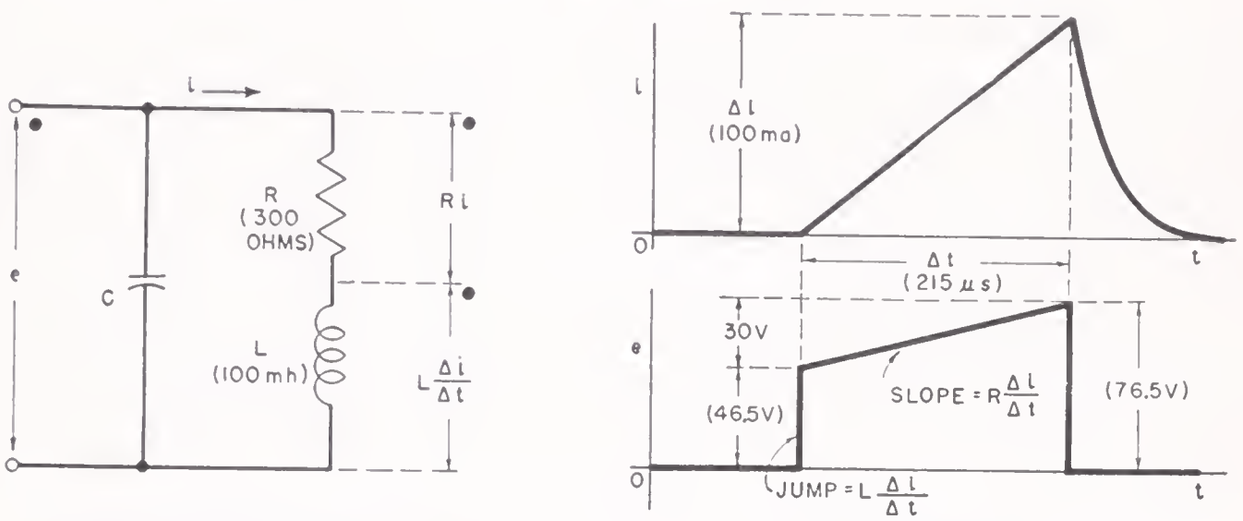
FIG. 26. Block diagram illustrating principles of magnetic cathode-ray-tube indicator.

to produce the desired beam deflection. Echo signals are applied to the control grid of magnetic cathode-ray tubes to control the intensity of the electron beam, and the sweep-current circuits must provide for the display of two coordinates of the objects located—for example, range and azimuth angle in Type B and Type P presentations.

13. Basic Magnetic-deflection Indicator. The block diagram of Fig. 26 illustrates the basic principles of an indicator employing a magnetic cathode-ray tube. Between sweeps, the electron beam is cutoff. The pulse from the multivibrator, which is initiated by a trigger pulse synchronized with the transmitted pulse, controls the range-sweep interval. During this interval, a saw-tooth current wave is sent through a set of deflection coils, and an intensifier pulse, applied to the cathode (or control grid, or first anode), causes the tube to operate at the threshold of illumination. Video pulses applied by the receiver to the control grid then cause bright spots to appear on the screen at points corresponding to the locations of objects. If range markers are used, they also are applied to the control grid.

If plan-position presentation is desired, the deflection coil is usually rotated in synchronism with the antenna. For Type B presentation, a second set of deflection coils at right angles to the first is provided. A current, the magnitude of which is determined by the direction in which the antenna is pointed during each range sweep, is sent through the second set of coils and provides the azimuth sweep. In some Type P indicators also, two sets of deflection coils are used, and the currents sent through them are of such a nature that a rotating deflection field is produced.

In the following articles, the circuits employed to produce linear range sweeps and to move the luminous spot in accord with antenna position are discussed. Means of obtaining plan-position indication with a rotating coil and with fixed coils are described.



(A) EQUIVALENT CIRCUIT (B) WAVEFORMS

FIG. 27. Equivalent circuit and current and voltage waves for deflection coil.

14. Current-Voltage Relations for Deflection Coil. An equivalent circuit for a deflection coil is drawn in Fig. 27*A*. The resistance and inductance of the winding are represented by R and L, and the distributed capacitance by an equivalent capacitor C connected across the coil. The voltage e applied to the coil causes a current i to flow through the turns of the coil. The magnetic field produced by this current deflects the electron beam. An additional current flows through the distributed capacitance while e is changing, but the capacitance current does not assist in deflecting the electron beam.

If a linear sweep is to be produced, the waveform of the current i should be a saw tooth, as in Fig. 27*B*. During the linear rise of current, the voltage across the deflection-coil inductance is constant (because di/dt is constant) and has the value $L\Delta i/\Delta t$, where Δi is the total change of current in the sweep interval Δt. The voltage across the coil resistance is Ri and has the same saw-tooth waveform as the current. The coil voltage e is the sum of the rectangular inductance-voltage pulse and the

triangular resistance-voltage pulse and has the trapezoidal waveform of Fig. 27*B*.

A simple relation exists between the jump and the sloping rise of the trapezoidal wave. The jump voltage is $L\Delta i/\Delta t$, and the slope is the rate of change of Ri which is $R\Delta i/\Delta t$. Therefore,

$$\frac{\text{Jump}}{\text{Slope}} = \frac{L\Delta i/\Delta t}{R\Delta i/\Delta t} = \frac{L}{R} = \text{time constant of coil} \tag{2}$$

The total rise of the sloping portion of the e wave is the product of the slope and Δt; therefore,

$$\frac{\text{Jump}}{\text{Sloping rise}} = \frac{\text{jump}}{\text{slope} \times \Delta t} = \frac{\text{jump/slope}}{\Delta t} = \frac{\text{time constant of coil}}{\text{time interval of sweep}} \tag{3}$$

Equation 3 indicates the relative importance of jump and slope. For a long-duration slow-moving sweep, the jump is negligible relative to the sloping rise. The deflection-coil voltage has nearly the same saw-tooth shape as the current. For a very fast sweep, the opposite is true; the sloping rise is small relative to the jump, and the deflection-coil voltage wave is nearly rectangular. For intermediate sweep speeds, both jump and slope are important.

In Fig. 27, typical deflection-coil parameters are indicated in parentheses: $R = 300$ ohms, $L = 100$ mh, and $\Delta i = 100$ ma for full-scale deflection of the electron beam. A sweep range of 20 statute miles, corresponding to

$$\Delta t = 10.74 \times 20 = 215\ \mu\text{s}$$

is assumed. The inductance voltage is

$$L\frac{\Delta i}{\Delta t} = 100 \times 10^{-3} \times \frac{100 \times 10^{-3}}{215 \times 10^{-6}} = 46.5 \text{ volts}$$

The rise of the sloping portion of the voltage wave is

$$R\Delta i = 300 \times 100 \times 10^{-3} = 30 \text{ volts}$$

and the peak value of e is the sum of $L\Delta i/\Delta t$ and $R\Delta i$, or 76.5 volts. The ratio of jump to sloping rise is $46.5/30 = 1.55$, which value checks Eq. 3:

$$\frac{L/R}{\Delta t} = \frac{333\ \mu\text{s}}{215\ \mu\text{s}} = 1.55$$

Because of the relatively large deflection-coil currents and voltages required (100 ma and 76.5 volts), a small power-amplifier tube (for example, the Type 6L6 beam power tube) must be used as a deflection-coil driver tube. Note that Eqs. 2 and 3 apply to the *deflection-coil voltage*, not to the grid voltage of the driver tube.

Capacitance C in Fig. 27*A* does not enter the relation between deflection-coil voltage e and the current i; an exactly linear current wave is obtained no matter what the size of C *if* the voltage wave e in Fig. 27*B* is impressed. However, capacitance C prevents any physical circuit from producing the sudden jump of voltage required at the beginning of the sweep. The leading edge of the trapezoid must slope slightly (the

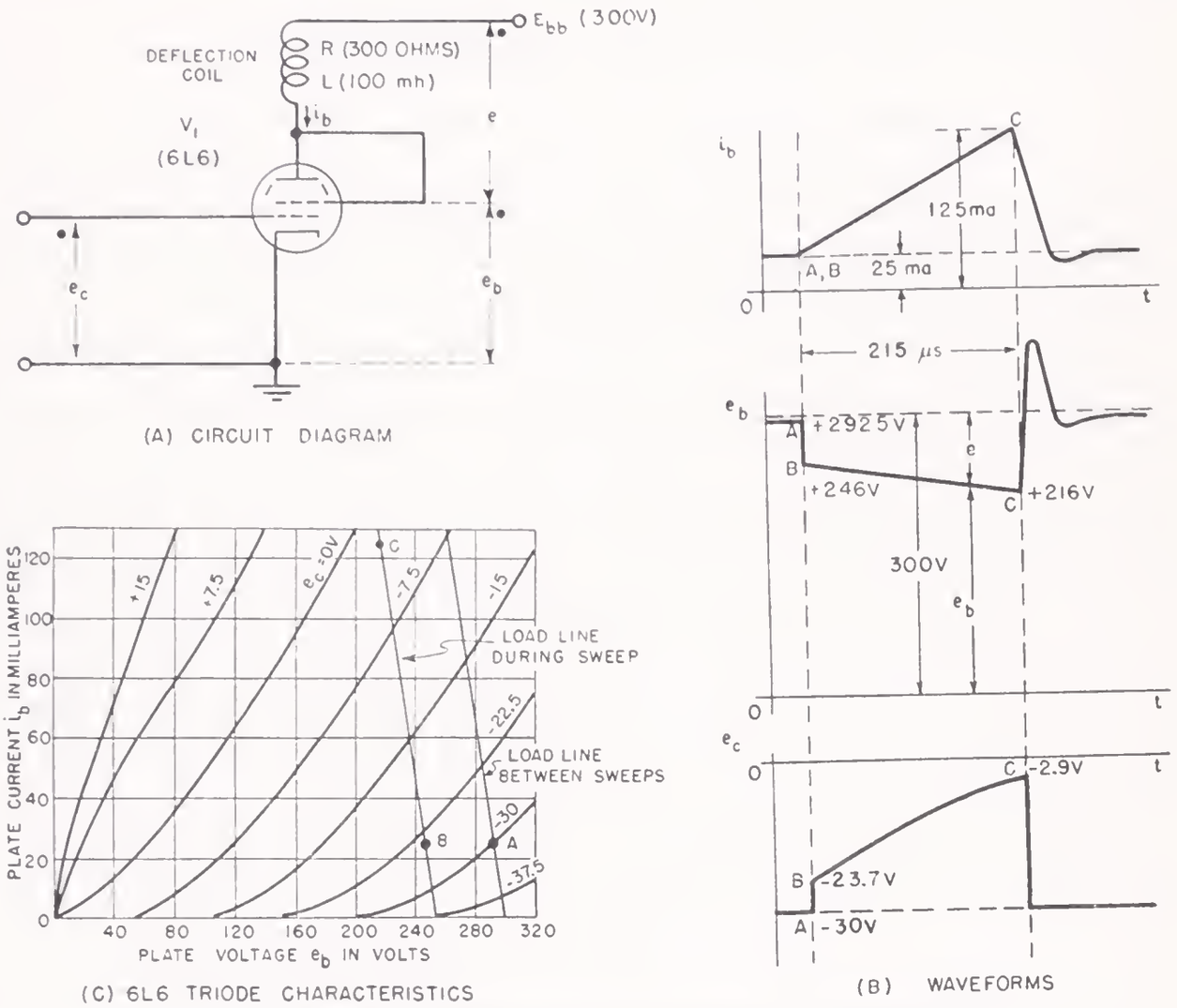


FIG. 28. Driver-tube voltage and current waveforms.

amount depending upon the size of C and the current supplied by the source of e), and as a result the sweep current is slow in starting. Minimizing the value of C and the internal resistance of the source of e is therefore of importance where fast sweeps requiring large jump voltages are desired.

15. Grid Voltage for Deflection-coil Driver Tube. A circuit for supplying current to a deflection coil is drawn in Fig. 28*A*, and waveforms in this circuit are drawn in Fig. 28*B*. For simplicity, the deflection-coil capacitance is assumed negligible, and the current i effective in deflecting the electron beam is taken to be the same as the plate current i_b. Thus the desired wave of plate current is a saw tooth, and the corresponding wave of e_b, which is the supply voltage decreased by the deflection-coil

voltage e_c is an inverted trapezoid. The grid-voltage waveform e_c required to produce these waves of i_b and e_b is to be determined.

The driver tube V_1 is assumed to be a triode-connected 6L6, and the deflection-coil constants and sweep duration of Art. 14 are used. The deflection-coil current is considered to rise from 25 to 125 ma (instead of from 0 to 100 ma) in order that the nonlinear low-current region of the 6L6 curves in Fig. 28*C* may be avoided.

Before the sweep interval begins, i_b is 25 ma and e_b is

$$300 - 25 \times 300 \times 10^{-3} = 292.5 \text{ volts}$$

Operation is at point A in Fig. 28*C*, and e_c is -30 volts. Point A may be located without computing e_b if a load line for $R = 300$ ohms and $E_{bb} = 300$ volts is drawn and the point for $i_b = 25$ ma found.

During the sweep e_b is less than 300 volts because of the inductance voltage $\frac{L\Delta i_b}{\Delta t} = 46.5$ volts and the resistance voltage Ri_b. The inductance voltage is equivalent to a battery opposing the plate supply voltage. Therefore, the operating point moves from point B to point C in Fig. 28*C* along a 300-ohm load line drawn 46.5 volts to the left of the line through point A. The initial and final values of e_c, determined from points B and C, are -23.7 and -2.9 volts. Intermediate points on the e_c wave may be located similarly. The e_c wave is trapezoidal, except for a slight curvature of the rise between points B and C caused by nonlinearity of the tube characteristics.

The proportions of the trapezoids in the plate-voltage and grid-voltage waveforms are different. In the e_b wave of Fig. 28*B*, the jump is somewhat greater than the sloping change; in the e_c wave, the jump is small relative to the sloping rise. This difference may be explained in terms of the linear equivalent circuit of Fig. 29, which is applicable for current and voltage changes and is approximately correct over the entire region of the tube curves used during the sweep. If Δi_b in this circuit is to be a saw-tooth waveform, Δe_b must be a trapezoidal voltage of jump-to-slope ratio L/R. Similarly, the generator voltage $\mu\Delta e_c$ must be a trapezoidal wave of jump-to-slope ratio $L/(R + r_p)$, and the grid-voltage wave must be a smaller trapezoid of these same proportions. To the extent that tube characteristics are linear, Eqs. 2 and 3 may be applied to determine the proportions of the driver-tube grid-voltage wave, provided the time

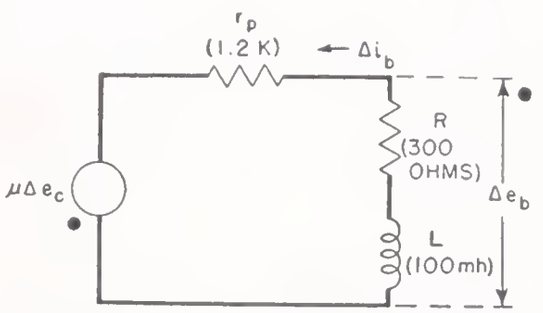


FIG. 29. Linear equivalent circuit for Fig. 28*A* applicable during sweep period.

constant of the *plate circuit* is used rather than that of the deflection coil alone.

In the example of Fig. 28, the average value of r_p during the sweep is 1.2 K, and the average value of μ is 7.5. The jump in the grid-voltage waveform is thus $(L\Delta i_b/\Delta t)/\mu$, or $46.5/7.5 = 6.2$ volts, and the sloping rise is $\Delta i_b(r_p + R)/\mu$, or

$$\frac{100(1.2 + 0.3)}{7.5} = 20 \text{ volts}$$

These values agree approximately with those obtained graphically and indicated in Fig. 28*B*.

A triode connection of the 6L6 was assumed in the above example in order to make the jump in the grid-voltage wave noticeable. More common practice is to employ the beam-tetrode connection. Then the plate resistance of the tube is high, and except for very short-range sweeps, the jump in the e_c wave is only a small portion of the total rise. For example, if the deflection coil of Fig. 28 is used with a tube having a plate resistance of 20 K, the plate-circuit time constant is $L/(R + r_p)$ or $100/(0.3 + 20) \approx 5$ μs; and for a 215-μs sweep, the ratio of jump to sloping rise is only $\frac{5}{215}$ or about 2 per cent. If a high-resistance driver tube is employed, the jump voltage is usually omitted from the e_c waveform for sweep ranges of about 40 miles or more. Often the waveforms associated with high-resistance driver tubes are saw-tooth grid voltage, nearly rectangular plate voltage, and saw-tooth plate current.

Wiring and deflection-coil capacitances thus far have been neglected. For relatively slow sweeps (sweep ranges greater than about 20 miles), these capacitances have little effect during the sweep interval, and the waveforms of Fig. 28*B* are applicable for this interval. If the grid voltage falls suddenly at the end of the sweep, the capacitances will affect the transient produced by the sudden decrease of plate current. A ringing oscillation in the deflection coil, damped by coil losses and the plate resistance of the tube, often occurs, as indicated in Fig. 28*B*. A 25- to 50-K resistor is sometimes connected across the deflection coil to provide additional damping and to limit the peak voltage produced. Use of a damping resistor is especially important if the driver tube is cutoff between sweeps. Because the damping resistor is effectively in parallel with the plate resistance of the tube, it increases the plate-circuit time constant during the sweep interval. Consequently, a higher jump-to-slope ratio for the grid-voltage trapezoid is required when a damping resistor is employed.

The effect of capacitance on the sweep waveform itself—a rounding of the front corner of the saw tooth caused by a sloping leading edge of the plate-voltage trapezoid—is of importance with high-speed sweeps.

Because of the capacitance, accurately linear sweeps for ranges of about 1 mile or less are very difficult to produce. One method of reducing the effect of shunt capacitance is to decrease the output resistance of the driver tube. To this end, a cathode follower is sometimes used to supply current to the deflection coil. The low resistance of the follower allows current to be supplied to the deflection coil and its capacitance with little distortion of the trapezoidal output voltage. Another method is to use the driver-tube circuit of Fig. 28*A* and to add a "spike" to the leading

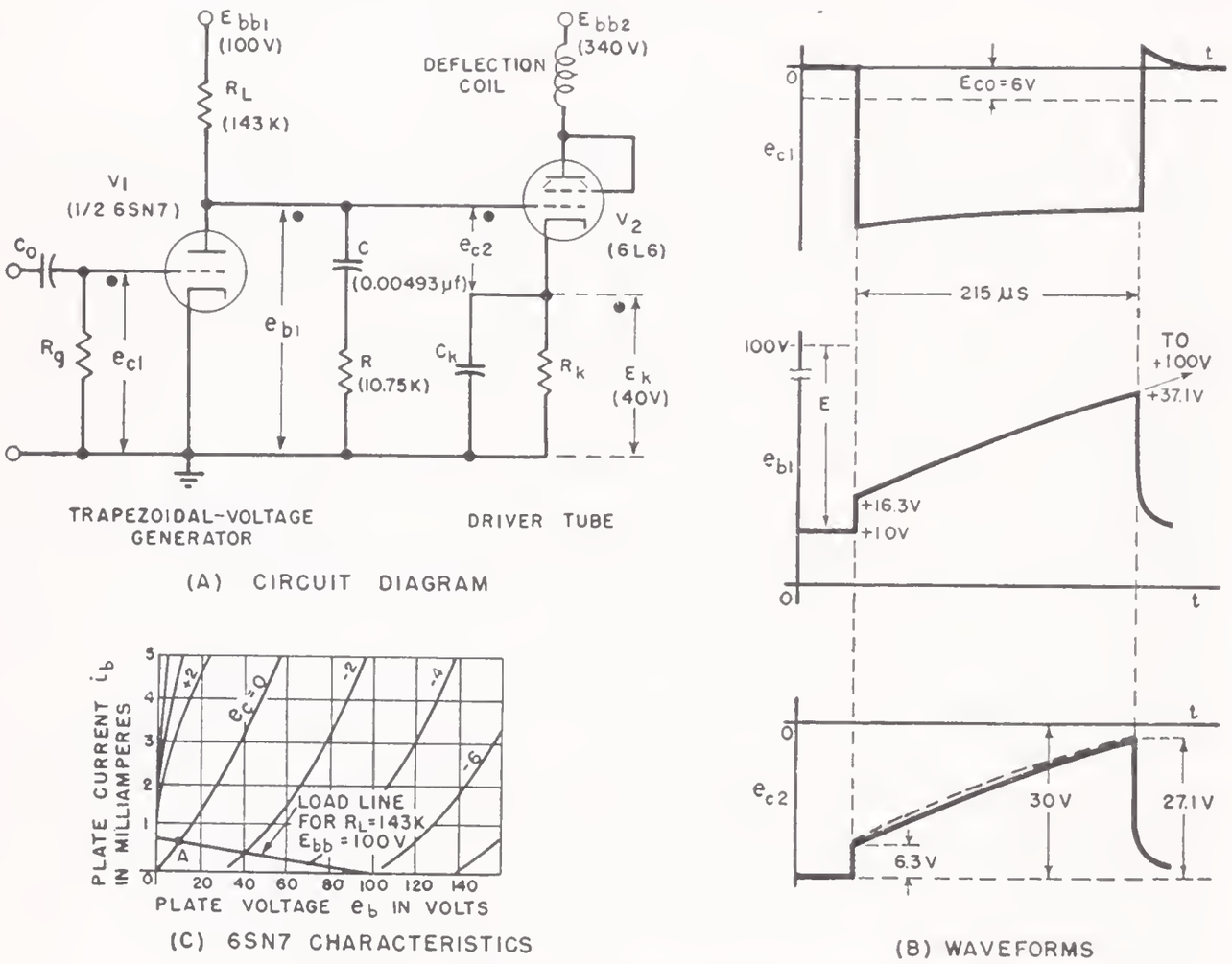


FIG. 30. *R-C* trapezoidal-voltage generator.

edge of the trapezoid in the grid-voltage waveform. The spike causes the tube to draw a sudden pulse of current from the deflection-coil capacitance, thus allowing the plate voltage to decrease very suddenly, as required by the jump of the plate-voltage trapezoid.

16. *R-C* Trapezoidal-voltage Generator. The grid-voltage waveform required to produce a saw-tooth current waveform in the plate circuit of the driver tube is often obtained by means of the trapezoidal-voltage-generator circuit shown in Fig. 30*A*. The operation of this circuit closely resembles that of a saw-tooth generator (Art. 5, Chap. II), but because of resistor R between capacitor C and ground, the voltage generated has the shape of a trapezoid rather than that of a saw tooth. To illustrate the operation of this circuit, suppose that the waveform required

at the grid of V_2 is the e_c waveform of Fig. 28*B*. Between sweeps, V_1 is conducting with $e_{c1} = 0$, and operation is at point *A* on the plate characteristics (Fig. 30*C*). From these curves, e_{b1} is 10 volts. To make e_{c2} equal -30 volts, as required, E_k is adjusted to $+40$ volts. The plate supply voltage E_{bb2} must then be 340 volts in order to keep the plate-to-cathode voltage of V_2 at the 300-volt value used in the circuit of Fig. 28*A*.

To start the sweep a negative gate pulse (Fig. 30*B*) is applied to the grid of V_1. At the instant V_1 is cut off, the voltage e_{b1} jumps from 10 volts to a value which is

$$(100 - 10)\frac{R}{143 + R}$$

(since the voltage across C cannot change instantaneously). Therefore, the initial jump at the grid of V_2 will have the desired value of 6.3 volts if R is made equal to 10.75 K. As capacitor C begins to charge, e_{b1} rises exponentially toward 100 volts with a time constant of $(R_L + R)C$. The time constant must be chosen so that the voltage across R_L decreases from its initial value of $100 - 16.3 = 83.7$ volts to $100 - 37.1 = 62.9$ volts in 215 μs. Therefore,

$$62.9 = 83.7\epsilon^{-215/T}$$

and

$$T = 753\ \mu\text{s}$$

To obtain this time constant, C must have a value of

$$\frac{753}{(143 + 10.75)10^3} = 0.00493\ \mu\text{f}$$

The waveform of e_{c2} is the solid curve, and the required waveform is the dashed curve. If a closer agreement of the two is desired, different values of E_{bb1} may be used, and new values of R_L or R and C determined to make the exponential rise of e_{b1} more nearly that of the dotted curve.

If E represents the difference in voltage between E_{bb1} and e_{b1} when $e_{c1} = 0$ (see waveform of e_{b1}, Fig. 30*B*), then the jump voltage is

$$E_{\text{jump}} = ER/(R_L + R)$$

and the initial slope of the exponential rise is

$$\frac{E - E_{\text{jump}}}{(R_L + R)C} = \frac{ER_L}{(R_L + R)^2C}$$

The jump-to-slope ratio is therefore $RC(R_L + R)/R_L$, or for the usual condition of R very small relative to R_L, the jump-to-slope ratio is very nearly RC. The jump-to-slope ratio of the grid-voltage trapezoid produced is RC, a time constant of the trapezoidal-voltage generator, and the jump-to-slope ratio of the e_c trapezoid desired is the time constant of the

plate circuit of the driver tube. To make possible a linear saw-tooth wave of current, these two time constants must be equal. Change of the supply voltage E_{bb1}, or of R_L (provided R_L is much greater than R), does not affect the jump-to-slope ratio of the e_c wave. Thus either E or R_L may be adjusted to change the rate of rise of the sweep current without effect upon linearity of the current rise.

Usually a multivibrator is employed to generate the negative rectangular pulse that controls conduction in tube V_1 of Fig. 30*A*. The positive pulse from the normally on tube of the multivibrator is sometimes peaked and the peaks added to the e_{c2} trapezoid of Fig. 30*B* to improve the start of the sweep (see Art. 15). Resistor R and a small capacitor connected between the plate of the multivibrator tube and the plate of V_1 function as an R-C peaker.

Another method of adding a spike to the grid-voltage wave makes use of a very small cathode by-pass capacitor, C_k in Fig. 30*A*. The initial jump of the trapezoidal-voltage wave applied to tube V_2 is fully effective in changing the grid-cathode voltage because C_k prevents a sudden change of the cathode voltage. After a very short time, C_k charges and allows the cathode voltage to rise. Thereafter changes in the trapezoidal voltage wave are only partially effective in changing the grid-cathode wave. A spike in the grid-cathode voltage wave is thus produced. The trapezoidal-voltage generator must be designed to produce a jump in the grid-to-ground voltage equal to the desired jump in the grid-to-cathode voltage, plus the amplitude of the spike produced by the cathode circuit.

17. Rectangular Grid-voltage Pulse for Driver Tube. In Art. 15, the jump required in the driver-tube grid-voltage trapezoid is shown to be negligible when the sweep duration is long and the plate resistance of the driver tube high. On the opposite extreme, if the sweep duration is short and the plate resistance low, the sloping rise of the trapezoid becomes negligible relative to the jump, and a satisfactory sweep current is obtained if a rectangular pulse is applied to the grid of the driver tube.

A circuit for producing a sweep current in this way is indicated in Fig. 31*A*. Assume first that switch S is closed. Between sweeps, tube V_1 is cutoff, as indicated by the e_c wave of Fig. 31*B*, and the deflection-coil current is zero. During the sweep, tube V_1 conducts with zero grid voltage and the plate voltage e_b falls to a low value. The voltage difference $E_{bb} - e_b$ applied across the inductance of the deflection coil causes a steady increase of the deflection-coil current i. At the end of the sweep, the plate current i_b is reduced suddenly to zero, and were it not for R_1 and diode V_2, a ringing oscillation would occur in the deflection coil. This oscillation could be damped by connecting a resistor across the coil, as mentioned in Art. 15. Use of the diode in series with the resistor prevents R_1 from by-passing current around the deflection coil

during the sweep, and thus a low value of R_1 may be used to provide adequate damping.

Switch S and inductor L_1 make possible a choice of sweep durations. A short-range sweep is obtained with S closed; a longer range sweep is produced if the duration of the positive pulse applied to the grid of V_1 is increased with S open. The voltage difference $E_{bb} - e_b$ is then applied across L and L_1 in series, and the deflection-coil current increases more slowly.

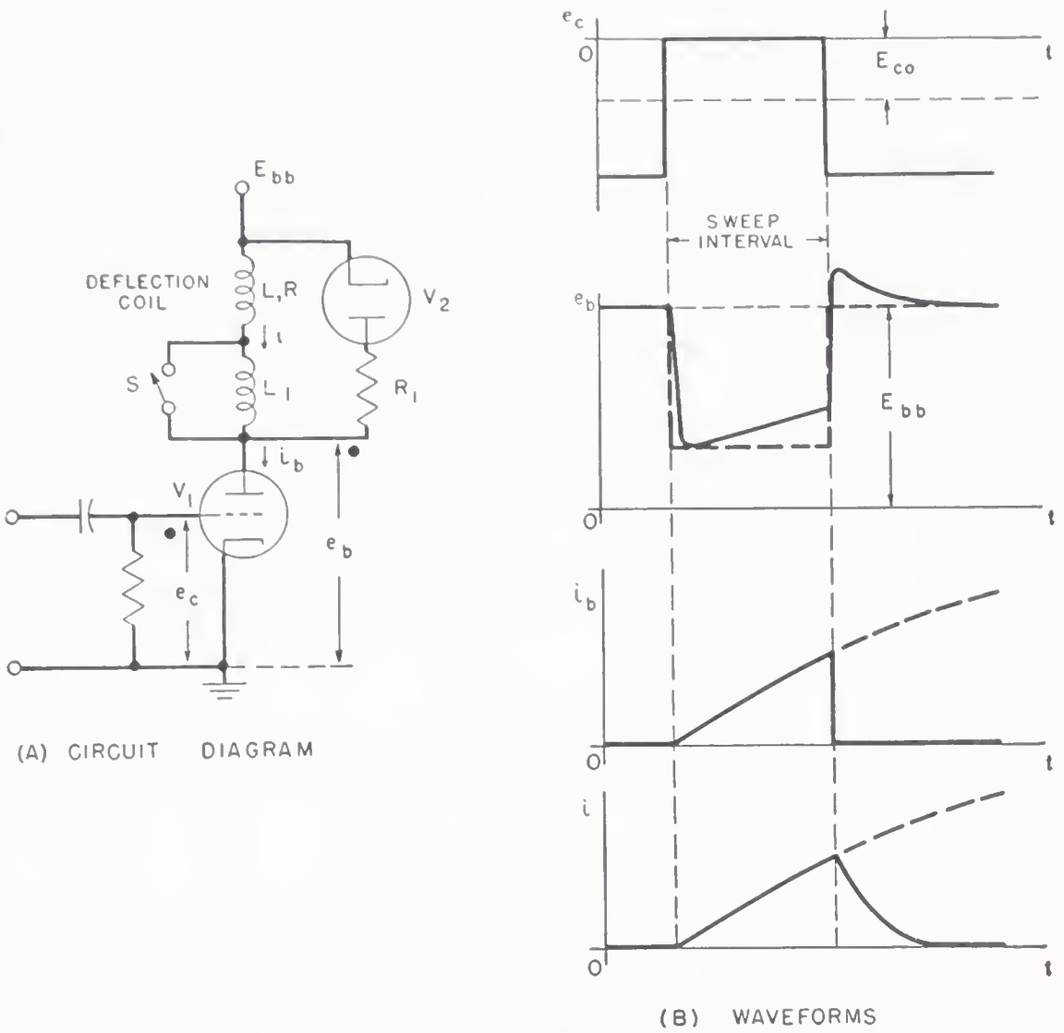


FIG. 31. Sweep circuit using rectangular grid-voltage pulse.

The conditions of operation during the sweep interval are simple if the plate resistance of tube V_1 may be considered negligible. Then the rectangular pulse applied to the grid of V_1 causes a rectangular e_b wave (see dashed waveforms in Fig. 31*B*), and the current i is the exponentially rising current caused by suddenly applying a steady voltage to an R-L circuit. The portion of the current rise used for the sweep is essentially linear if the final current value approached by the exponential ($E_{bb} - e_b$ divided by the circuit resistance) is many times larger than the maximum sweep current.

Ordinarily the driver-tube plate resistance cannot be ignored. The resistance has two effects: (1) The plate voltage e_b increases with increase

of current during the sweep. The final value approached by the exponentially rising current is therefore reduced and the sweep linearity impaired. In terms of time constants with switch S closed, the time constant of the exponential is reduced from L/R to $L/(R + r_p)$ because of the plate resistance r_p, and thus a greater portion of the exponential is used by a sweep of specified duration. (2) The decrease of e_b at the initial instant of the sweep is not instantaneous because of deflection-coil capacitance. Therefore, a rounded corner is produced at the beginning of the current saw

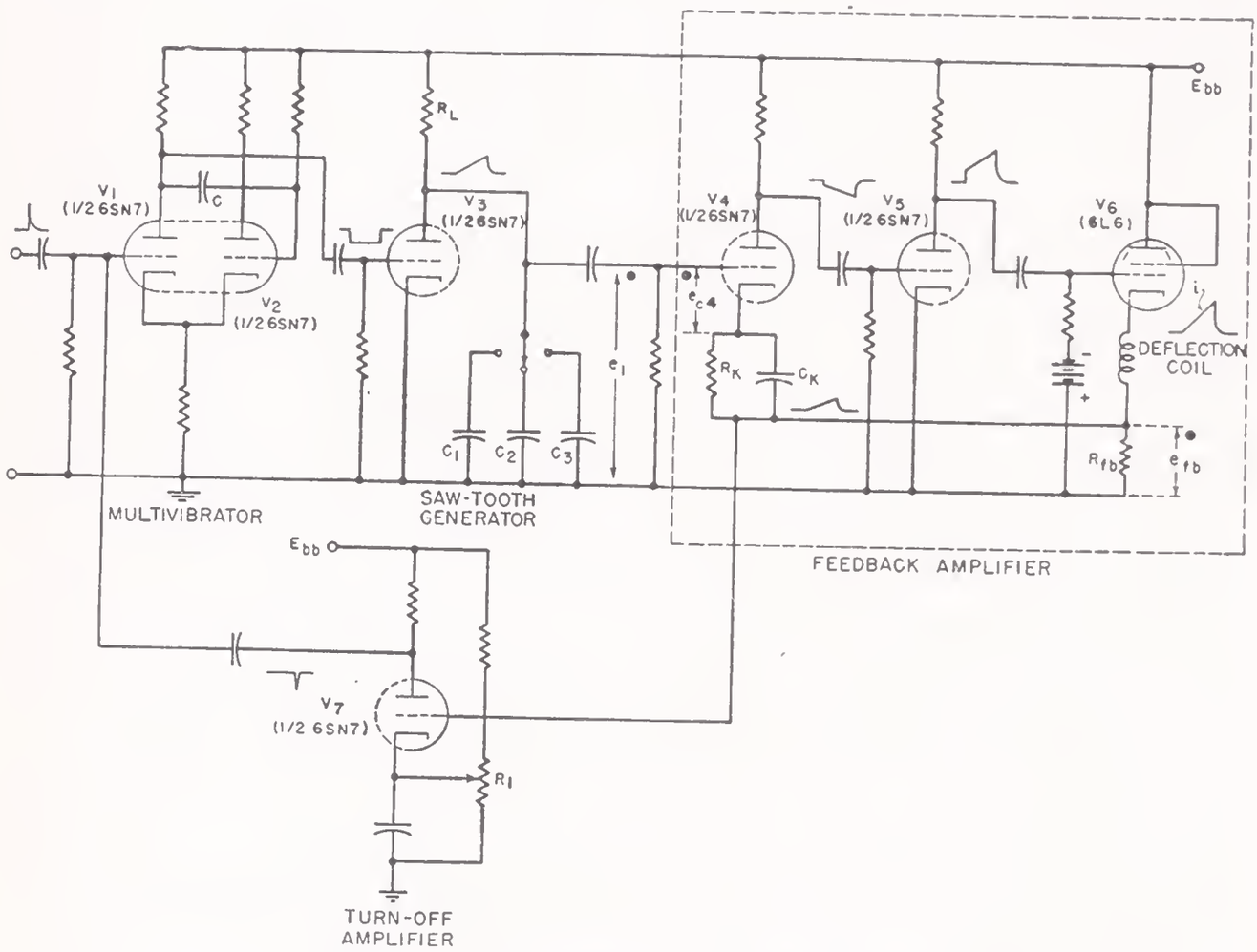


FIG. 32. Feedback-amplifier sweep circuit.

tooth. The solid e_b curve of Fig. 31B illustrates the waveform obtained when driver-tube plate resistance is not negligible.

18. Feedback-amplifier Sweep Circuit. A major cause of sweep-current nonlinearity, especially if the sweep duration is long enough so that capacitance effects are unimportant, is the nonlinear nature of amplifier-tube characteristics. Nonlinearities of this kind are greatly reduced by using a feedback amplifier as in Fig. 32. The feedback amplifier has the additional advantage that a saw-tooth waveform of input voltage may be used, a trapezoidal voltage of correct proportions to apply to the deflection coil being generated within the amplifier.

The circuit of Fig. 32 contains a multivibrator, a saw-tooth generator, and a three-stage feedback amplifier. The multivibrator produces a nega-

tive gate pulse that operates the saw-tooth generator, and the feedback amplifier converts the saw-tooth voltage wave applied to it into a saw-tooth current wave in the deflection coil.

Tubes V_4, V_5, and V_6 operate as linear amplifiers, and a voltage proportional to the output current of V_6 is fed back to the cathode of V_4 through resistor R_{fb}. The feedback is negative because a rise in voltage at the grid of V_4 causes a drop in voltage at the grid of V_5, a rise at the grid of V_6, a rise across both the deflection coil and R_{fb}, and a rise at the cathode of V_4.

The effect of the negative feedback is to make the deflection-coil current have nearly the same waveform as the voltage e_1. Tubes V_4, V_5, and V_6 amplify the grid-to-cathode voltage of V_4 about 500 times in producing the feedback voltage e_{fb}. Therefore, if the waveform of e_{fb} deviates by

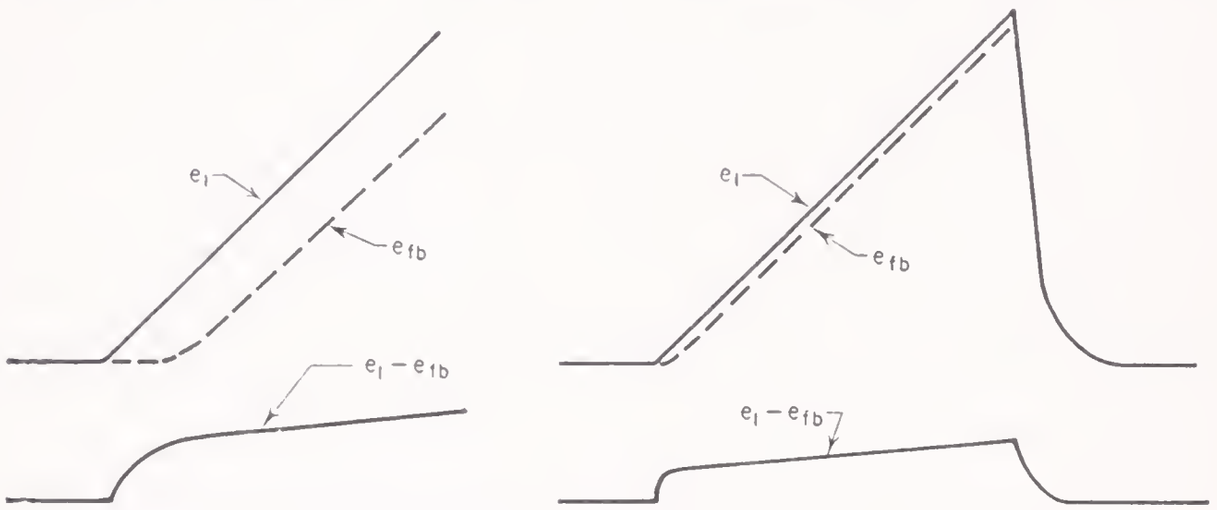

e1
efb
e1 − efb
e1
efb
e1 − efb
(A) EXPANDED-SCALE DRAWING SHOWING FIRST PART OF SWEEP
(B) NORMAL-SCALE DRAWING OF WAVEFORMS e1 AND efb (VERTICAL SCALE OF e1 − efb IS EXAGGERATED)

FIG. 33. Formation of trapezoid at input to feedback amplifier.

only a small amount from the saw-tooth waveform of e_1, the deviation is sufficient to produce a voltage e_{c4} which, when amplified, makes the waveforms of e_{fb} and e_1 similar. Thus the voltage across R_{fb} and, consequently, the current through it have similar saw-tooth waveforms.

The deflection-coil current, which is the same as the current in R_{fb}, has a saw-tooth waveform; therefore, the deflection-coil voltage must be trapezoidal. The grid voltage for V_6 must also be trapezoidal, and the grid voltage of V_5 and grid-to-cathode voltage of V_4 must be trapezoids of smaller amplitude. Figure 33*A* indicates how the nearly saw-tooth waveform of e_{fb} can be subtracted from the saw-tooth waveform of e_1 to produce the required trapezoid of e_{c4}. Note that a slight delay in the start of the linear rise of e_{fb} occurs, with the result that the sweep current is nonlinear at the very start of the sweep. The duration of the nonlinear part is about 1 μs in a typical circuit.

The amplifier circuit of Fig. 32 has negative *current* feedback, because the returned voltage is proportional to the load current. Negative cur-

rent feedback causes the output resistance of the amplifier to be high. Therefore, the waveform of the required input voltage e_1 is determined primarily by the waveform of the voltage drop across the amplifier output resistance and only to a small extent by the deflection-coil voltage requirements. Thus, even for short-range sweeps, the jump required in e_1 is so small compared with the sloping rise that the waveform is essentially a saw tooth.

Because the grid voltage of tube V_6 is never positive, the trapezoidal-sweep voltage at the grid of V_6 must have an average value equal to the grid-bias voltage. Consequently, when the sweep duration is changed to provide a different range scale, the between-sweeps value of the grid-to-cathode voltage must change (assuming that the repetition frequency remains constant). If V_6 is operated above cutoff between sweeps, the current through the deflection coil and therefore the starting position of the spot on the screen depend upon the setting of the range-selector switch. To ensure that the spot always starts from the same point regardless of the range scale, a clamper tube (Art. 10, Chap. II) is connected across the grid-return resistor.

In some circuits the grid-to-cathode voltage of V_6 is maintained below cutoff between sweeps to make the sweep always start at the same point. The grid voltage of V_6 must be raised rapidly to bring the tube into the linear operating region. The process of raising the grid voltage of V_6 from below cutoff to the point where the deflection-coil current begins to increase linearly causes a slight delay in the start of the sweep. Sometimes a compensating delay of the echo signals is provided by means of a delay line in the video section of the receiver.

In the circuit of Fig. 32, only the sweep capacitors need be switched to change range scales. Instead of switching capacitors in the multivibrator to control the sweep duration, a turn-off amplifier is used to control the width of the multivibrator pulse. Capacitor C in the multivibrator is so large that the multivibrator itself would produce a pulse of greater duration than that of the longest sweep used. The cathode voltage of V_7 is set by potentiometer R_1 so that the turn-off amplifier is normally cutoff. When the deflection-coil current reaches a predetermined value, the voltage e_{fb} becomes large enough to cause V_7 to draw current. As a result, a negative pulse is produced at the plate of V_7 which causes the multivibrator to switch back to its normal between-sweeps condition. The amplitude of the sweep trace on the cathode-ray tube is therefore controlled by adjustment of R_1 and is automatically maintained the same on all ranges. The sweep speed can be adjusted by varying the sweep capacitor C_1, C_2, and C_3 or the charging resistor R_L. Adjusting R_{fb} changes the sweep speed and trace length simultaneously in such a way that the sweep duration remains nearly constant.

Three shunt capacitances in the circuit of Fig. 32 are of relatively large size. One is the deflection-coil capacitance, and the other two are the capacitances between each deflection-coil terminal and ground. These last two are large because a cable is ordinarily required to connect the deflection coil to the amplifier. The capacitance between the cathode of V_6 and ground is unimportant because V_6, a cathode follower, can charge it very quickly. The other two capacitances affect considerably the start of the sweep because they may cause the deflection-coil current to differ from the current in R_{fb}, and the circuit acts to produce a saw-tooth current in R_{fb}. Deflection-coil capacitance slows the start of the sweep as in other circuits, but its effect is somewhat offset by the capacitance across R_{fb} which decreases the feedback voltage at the beginning of the sweep.

19. Azimuth-sweep Circuit for Type B Presentation. Type B presentation requires not only a vertical motion of the spot on the indicator screen to indicate range but also a horizontal motion of the spot to indicate the azimuth angle of reflecting objects. A variety of methods are used for correlating the horizontal motion of the spot with the azimuth angle of the antenna beam. One method, often used when azimuth scan is obtained by mechanical motion of the antenna system, employs a potentiometer geared to turn with the antenna. The potentiometer supplies current to the deflection coil, and thus the spot deflection is made a function of the direction in which the antenna radiates maximum energy.

Frequently, the antenna is capable of rotating continuously, but only a sector of 180 deg or less from the 360-deg region scanned is presented on the Type B indicator. A potentiometer having a continuous circular winding as in Fig. 34*A* is then used. A direct voltage is applied to fixed taps on the winding, and the deflection coil is connected to the winding through two sliding contacts. The rotating member which carries the contacts is geared to turn at the same speed as the antenna. Angle θ in Fig. 34*A* is therefore the azimuth angle of the antenna.

Assume that the sector to be presented on the indicator lies between $\theta = -90$ deg and $\theta = +90$ deg. When θ is zero, the two sliding contacts are at points of equal potential, the deflection-coil current is zero, and the range trace lies in the center of the indicator screen. As θ changes from -90 deg through zero to $+90$ deg, the deflection-coil current varies from a maximum in one direction to a maximum in the other direction, as indicated in Fig. 34*B*. Consequently, the range trace moves from one side of the screen through the center position to the opposite side. Rotation of the antenna through the 180-deg sector beyond $+90$ deg, moves the range trace in the reverse direction across the indicator screen. Because objects in this sector are not to be presented, the indicator screen

is blanked in this part of the antenna motion. A cam-operated switch controls the supply of voltage to an electrode of the cathode-ray tube.

Because the rate of change of deflection-coil current corresponding to the antenna motion is very low, the voltage produced across the deflection-coil inductance is negligible. Therefore, the coil is represented by its resistance, as in Fig. 34*A*, instead of by the complete equivalent circuit of Fig. 27*A*, p. 253.

Figure 34*B* shows the relation between deflection-coil current and azimuth angle to be linear. A linear relation is desired so that the azimuth scale of the Type B presentation may be linear. Unfortunately, because of the relatively large current required by the deflection coil, the

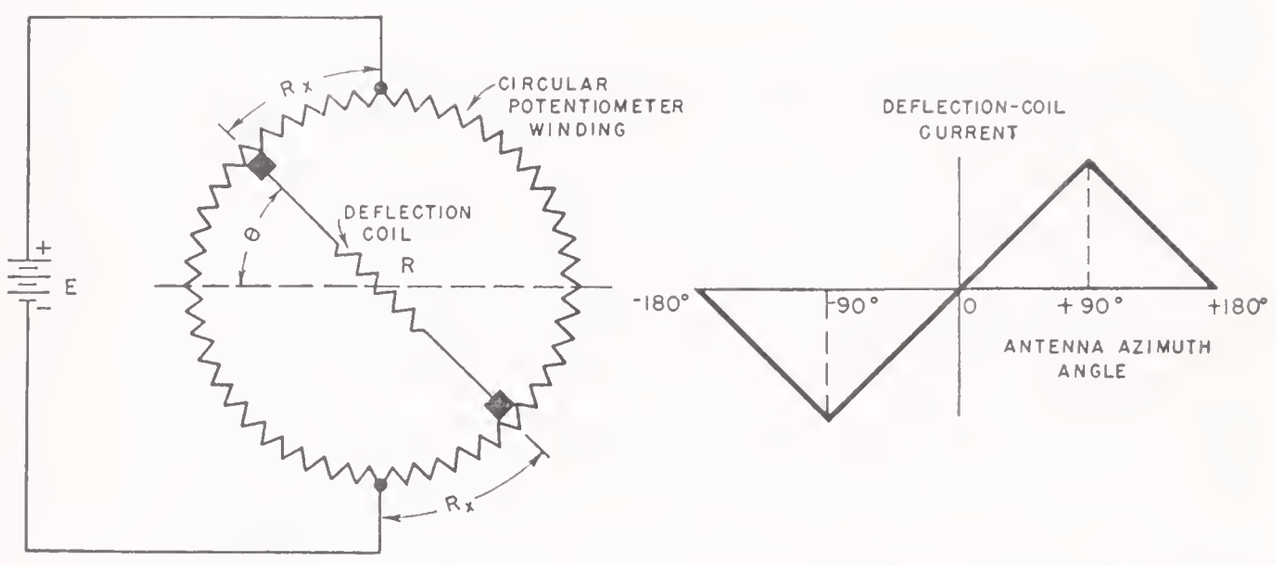


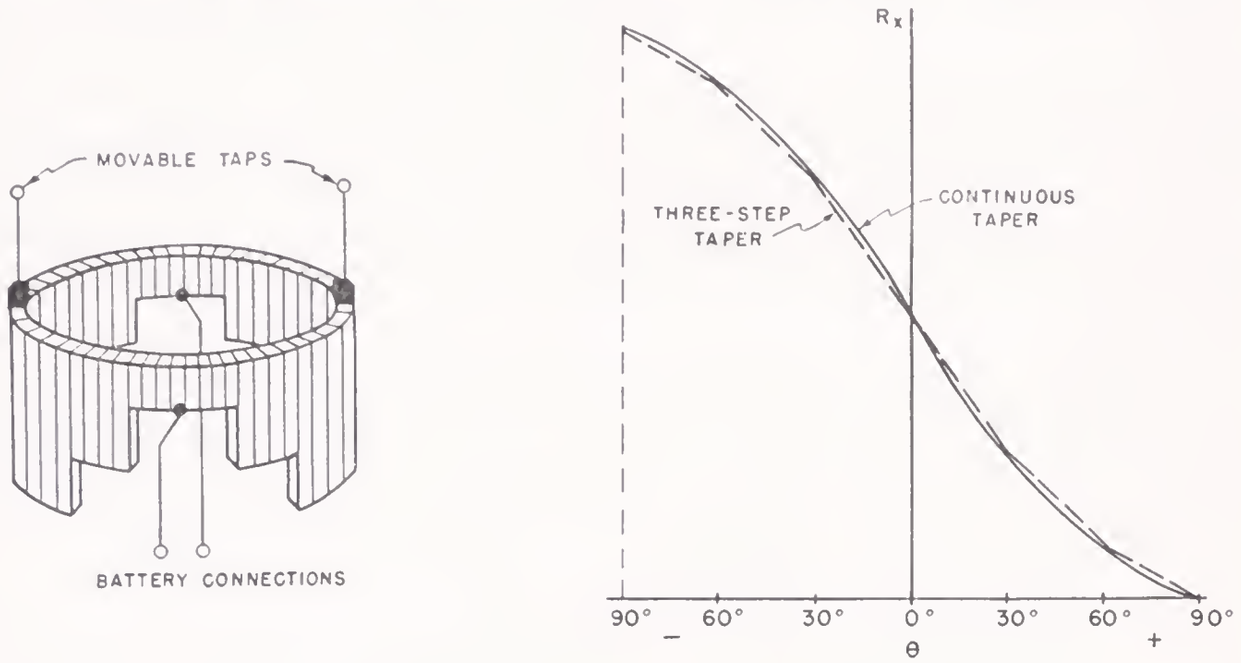
FIG. 34. Azimuth-potentiometer circuit and relation of deflection-coil current to azimuth angle.

linear relation of Fig. 34*B* would not be obtained if the potentiometer winding were uniform. With a uniform winding, resistance R_x in Fig. 34*A* would decrease linearly with θ as θ increased from zero to 90 deg. The deflection-coil voltage would therefore vary linearly only if the current in R_x were constant. Actually, the current in R_x increases as θ increases. Therefore, the resistance per unit angle of the winding must be less near the battery taps than it is midway between the taps; a *tapered* potentiometer winding is required to produce a linear current-angle relation.

A common construction of azimuth potentiometer is illustrated in Fig. 35*A*, and the resistance-angle relation for it is shown in Fig. 35*B*. The solid-line curve of Fig. 35*B* is the desired taper. (The exact shape required depends upon the relation of deflection-coil resistance to total potentiometer resistance.) The potentiometer is wound with wire of uniform size, and the cross section of the winding form is varied in three

steps. The broken line of Fig. 35*B* is the actual curve for the potentiometer; it approximates the desired smooth curve.

Among the other methods of providing azimuth sweep for the Type B indicators are the following: (1) A uniformly wound potentiometer may be used to supply voltage to an amplifier, the amplifier supplying the deflection-coil current. In a similar system used with an electrostatic Type B indicator tube, the amplifier supplies voltage to the deflection plates. (2) A synchro (see Sec. A, Chap. V, especially Art. 12) may be geared to the antenna and made to furnish an alternating voltage of magnitude dependent on antenna position. A phase-sensitive rectifier (see Art. 16, Chap. V) changes the synchro output to direct voltage, and a d-c



(A) CONSTRUCTION OF POTENTIOMETER (B) VARIATION OF R_x WITH θ

FIG. 35. Tapered azimuth potentiometer.

amplifier supplies the deflection-coil current. (3) In rapid-scan systems, the antenna beam rotates or moves back and forth several times a second. Often a small a-c generator operating in synchronism with the scan furnishes the deflection-coil current either directly or through an amplifier.

Several presentations other than Type B (for example, Types C, D, and H) require azimuth sweeps. These are obtained in much the same way as the azimuth sweep for Type B. Some presentations (for example, Types C, D, and E) require elevation sweeps. The same basic methods are used again, but of course the deflection-coil current is made to depend upon the elevation angle of the antenna beam instead of upon the azimuth angle.

20. Rotating-coil Plan-position Indicator. For Type P presentation, range-sweep circuits of the types described in Art. 13 through 18 are employed, and the range trace is made to begin at the center of the screen. Type P azimuth indication requires that the range trace rotate

about the center of the screen. When antenna scan is produced by rotation of the antenna, a very simple means of providing plan-position azimuth indication is to cause the deflection coil to rotate about the neck of the cathode-ray tube in correspondence with the antenna motion.

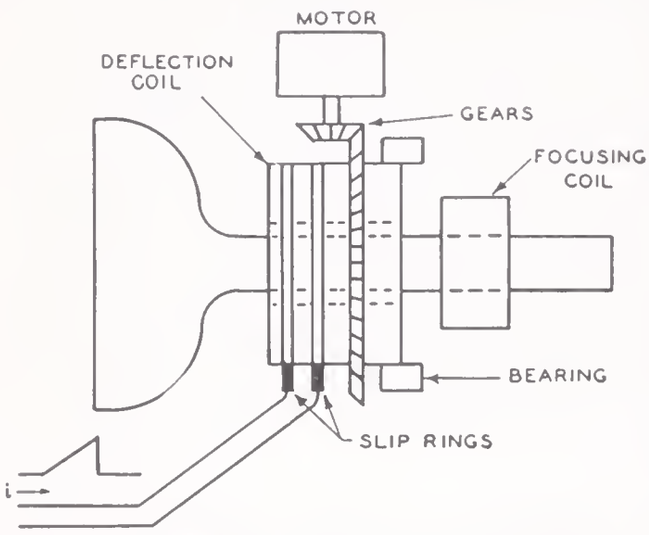


FIG. 36. Cathode-ray tube and rotating deflection coil used in plan-position indicator.

Figure 36 shows the tube and rotating coil. The yoke carrying the deflection coil is mounted in bearings and driven by a motor. Slip rings are provided so that range-sweep currents may be sent through the coil. Figure 37 illustrates the way in which azimuth indication is provided. The magnetic field established by range-sweep currents rotates with the deflection coil, and the sweep trace is always in a direction perpendicular to this field.

A number of different methods are used for rotating the deflection coil in correspondence with the antenna. If the antenna rotates continuously and at a constant rate, a simple method is to drive antenna and

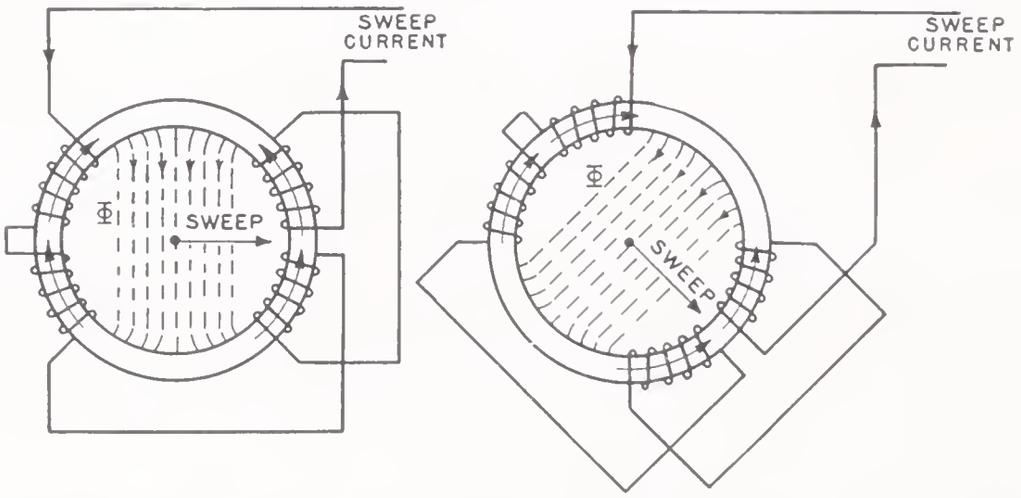


FIG. 37. Relation of sweep direction to deflection-coil position. (The electron beam is directed into the paper.)

deflection coil by synchronous motors connected to the same power supply. The coil then turns at exactly the same speed as the antenna, but it is possible for the two rotating systems to be differently oriented. To avoid errors from this source, a cam-operated switch on the antenna mount is arranged to brighten the range trace each time the antenna passes the zero-azimuth position. Thus a bright line should appear at an azimuth angle of zero on the indicator. If the line appears in a different position, power to the deflection-coil motor may be interrupted

momentarily to allow the coil to slip back in position relative to the antenna until correct alignment is obtained.

More common methods of driving the deflection coil employ synchros and servomechanisms. These methods, which permit the antenna to turn back and forth as well as to rotate continuously, are discussed in detail in Chap. V. Some of the possibilities are:

1. The antenna may be rotated by hand or by a motor, a synchro generator geared to turn with the antenna, and a synchro motor employed to drive the deflection coil. The two synchros are connected electrically, and the motor turns in correspondence with the generator. Small errors between the positions of motor and generator exist. To reduce the effects of these errors, the synchros are usually geared to turn at a higher speed than the antenna and the coil.

2. The antenna may be driven and a low-power servomechanism used instead of the synchros to keep the coil in step with the antenna. The servomechanism usually includes a synchro generator at the antenna, a synchro control transformer at the coil, and a motor to drive the coil in such a way as to minimize the difference between the positions of the generator and control transformer.

3. The deflection coil may be driven by hand or by a small motor. A higher power servomechanism is then used to drive the antenna in correspondence with the deflection coil.

21. Methods of Positioning the Sweep Trace. Articles 15 through 18 describe circuits for producing a current of the saw-tooth waveform required for a range sweep. In these articles, only the change of current required to produce the sweep, or the a-c component, is considered. The d-c component also is of importance because it determines the position of the trace on the indicator screen. Positioning requirements for Type B and Type P presentations are different. For Type B, the deflection-coil current must have a negative value at the beginning of the sweep to deflect the spot to the bottom of the screen. During the sweep, the current rises to zero (the spot is then at the center of the screen) and on to a positive value equal in magnitude to the initial negative value. For Type P presentation, the deflection-coil current must start at zero and increase sufficiently to move the spot from the center nearly to the edge of the screen.

Provision for positioning the range trace may be made in many different ways. Two circuits sometimes used with Type B indicators are shown in Fig. 38. The circuit in Fig. 38*A* employs *L*-*C* coupling to remove the direct component of plate current from the deflection-coil circuit. Inductances L_1 and L_2 and capacitor C are very large, so that the alternating components of deflection-coil current and voltage are essentially the same as the a-c components of plate current and voltage. Thus the

trapezoidal voltage e_{cn} produces a saw-tooth wave of current i in the same way as in the simpler circuit of Art. 15. If the E, R, L_2 branch of the circuit were open, the average current through the deflection coil would be zero because of capacitor C. The dotted line on the i wave of Fig. 38*A* would then be the zero axis for the wave, and the range trace would extend farther above the center of the tube than below. To secure correct positioning, a d-c component must be sent through the deflection-coil in the direction opposite to the reference arrow of i. Supplying this current is the function of the E, R, L_2 branch. Resistor R is made adjustable to serve as a positioning control.

The circuit of Fig. 38*B* employs a push-pull amplifier and a split-winding deflection coil. Connections are made so that current in L_1 in the

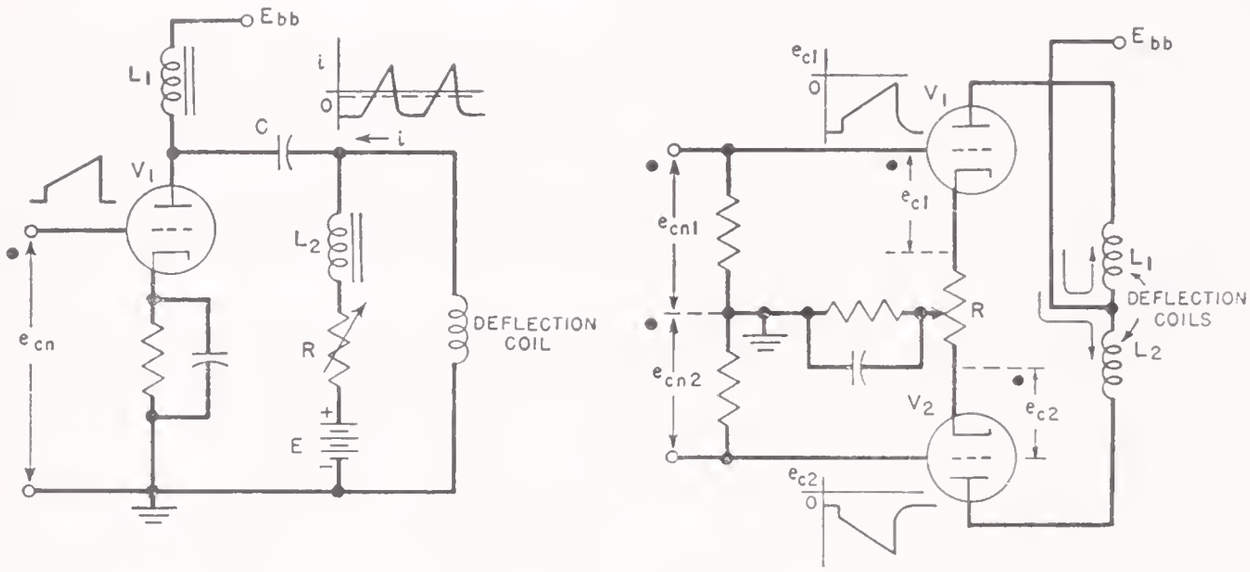


FIG. 38. Circuits that provide correctly positioned sweeps for Type B indicators.

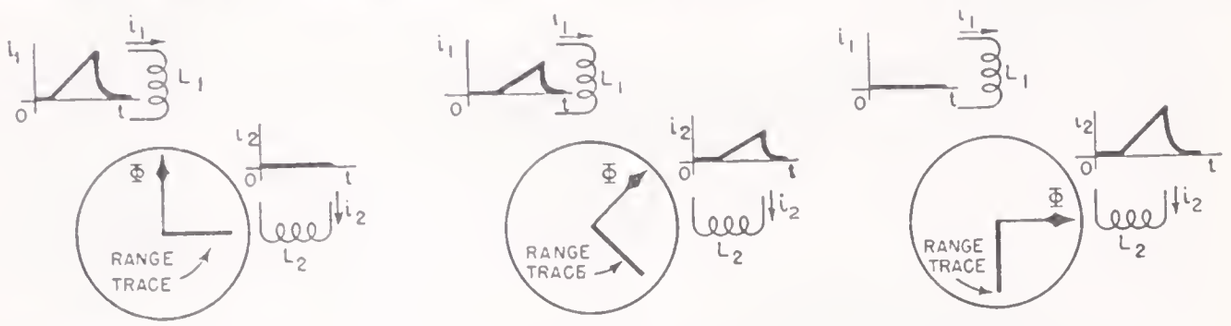
direction of the arrow moves the indicator spot upward and current in L_2 in the direction of the arrow moves the spot downward. At the beginning of the sweep (note the grid-voltage waveforms), current in V_1 and L_1 is small relative to that in V_2 and L_2, and the spot is near the bottom of the indicator screen. During the sweep, current in V_1 increases and that in V_2 decreases, causing the upward spot motion. Resistor R serves to control the division of current between V_1 and V_2 and thus to position the range trace.

A third positioning method for a Type B indicator makes use of an auxiliary deflection yoke and coil to which is supplied a steady current. The saw-tooth wave applied to the main deflection coil may then have any reasonable direct-current component, the current in the auxiliary coil is adjusted to secure correct positioning.

The range trace for a plan-position indicator is correctly positioned if the deflection coil is in series with the driver-tube plate circuit and the driver tube is cutoff between sweeps (as it may be, for example, in the

feedback circuit of Art. 18). The sweep current then starts from zero, and the range trace from the center of the screen. With circuits like the one in Art. 15, the low-current region of the driver-tube characteristics cannot be used because of nonlinearity, and a coupling circuit (for example, the circuit of Fig. 38.1) must be employed. Occasionally, where objects at very short range are important, the Type P sweep current is made to begin with a small positive value. Then the spot that normally indicates zero range is changed into a small circle about the center of the tube. Such a sweep causes a little distortion of the maplike presentation but permits better separation of near-by objects.

Where Type P presentation of only a relatively small sector is desired, off-center plan-position indication is often employed. The zero-range point is moved nearly to the edge of the indicator screen, and a wedge-shaped display of the desired sector is obtained. Such a display requires



(A) CURRENT IN L_1 ONLY (B) CURRENT IN L_1 AND L_2 (C) CURRENT IN L_2 ONLY

FIG. 39. Method of producing rotating range trace with fixed deflection yoke.

the use of two deflection coils on separate yokes. The main coil carries the normal range-sweep current and rotates with the antenna, the cathode-ray tube being blanked when regions outside the desired sector are scanned. The auxiliary deflection coil is stationary and carries a steady current to provide the off-center deflection for the display.

22. Fixed-coil Plan-position Indicator. In some plan-position indicators, instead of a rotating deflection coil special circuits are employed with fixed coils to provide the rotating magnetic field and range trace. The fixed deflection yoke (which is the same as the yokes used in Type B indicators and is illustrated in Fig. 5*D*, p. 224) has two sets of deflection coils, one for producing horizontal deflection of the luminous spot and the other for vertical deflection. Range-sweep currents are applied to the two sets of coils simultaneously, and the direction of the trace obtained depends upon the relative amplitudes of the two currents.

Figure 39 illustrates the method of producing the rotating range trace. In Fig. 39*A*, a range-sweep current is applied to deflection coil L_1 only, and the magnetic field lies along the axis of L_1. The resulting range trace is horizontal because the electron beam is deflected in a direction perpendicular to the magnetic field. In Fig. 39*B*, range-sweep currents are

applied to both coils, and the resultant magnetic field takes a position between the two coil axes. Because of this shift of the magnetic field, the range trace is rotated clockwise from its original position. The two current waves must be exactly alike and must be applied simultaneously; otherwise the trace would not be a straight line. In Fig. 39*C*, sweep current is applied to coil L_2 only, and the range trace lies 90 deg clockwise from its first position. The range trace may be made to turn smoothly through an angle of 90 deg if, during a large number of sweep intervals, the current applied to coil L_1 is slowly decreased to zero and the current in L_2 is simultaneously increased from zero. Further rotation is obtained if the deflection-coil currents are reversed in proper sequence.

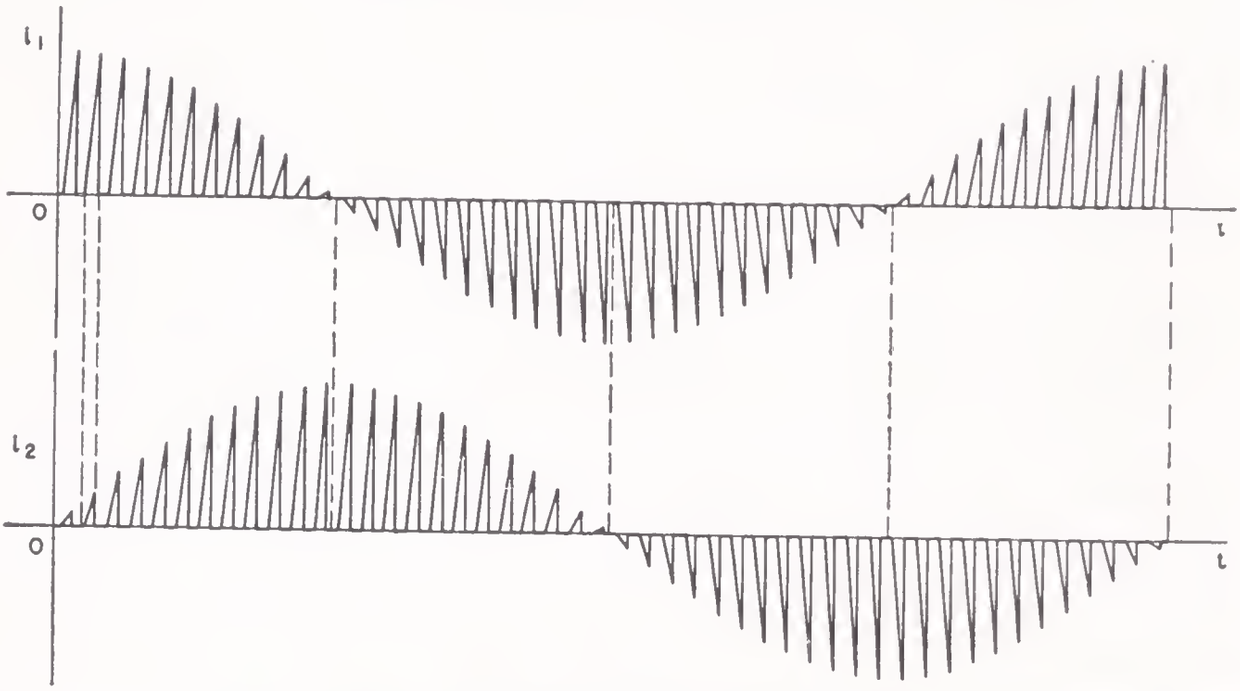


FIG. 40. Current in deflection coils of Fig. 39 required to produce rotating range trace.

In Art. 10 it is shown that a horizontal and a vertical displacement vector may be added together to produce a rotating vector. The resultant vector is of constant length and rotates with constant angular speed if the components execute sinusoidal variations equal in amplitude and 90 deg apart in phase. This result may be applied to the sweep trace in Fig. 39. In order that the range trace have a constant length and rotate at constant speed, it is necessary that the envelopes of the saw-tooth sweep currents be sinusoidal, as indicated by the waveforms of Fig. 40. The two envelopes differ in phase by 90 deg and have equal peak values if the two deflection coils have equal deflection sensitivities. Note that the individual saw-tooth waves of current are applied to the two coils *simultaneously;* the phase difference of 90 deg applies only to the variation of amplitude from one saw tooth to the next.

One of the many circuits used to produce sweep-current waveforms of the form of Fig. 40 is illustrated in Fig. 41. Because of the need for

reversing the directions of the horizontal and vertical components of magnetic field, each of the deflection coils of Fig. 39 is replaced in Fig. 41 by two coils connected to push-pull driver tubes. The four driver tubes require four trapezoidal grid-voltage waves. The amplitudes of the trapezoids must vary sinusoidally with antenna angle, as shown in Fig. 42, in order that sweep current having a similar variation may be obtained.

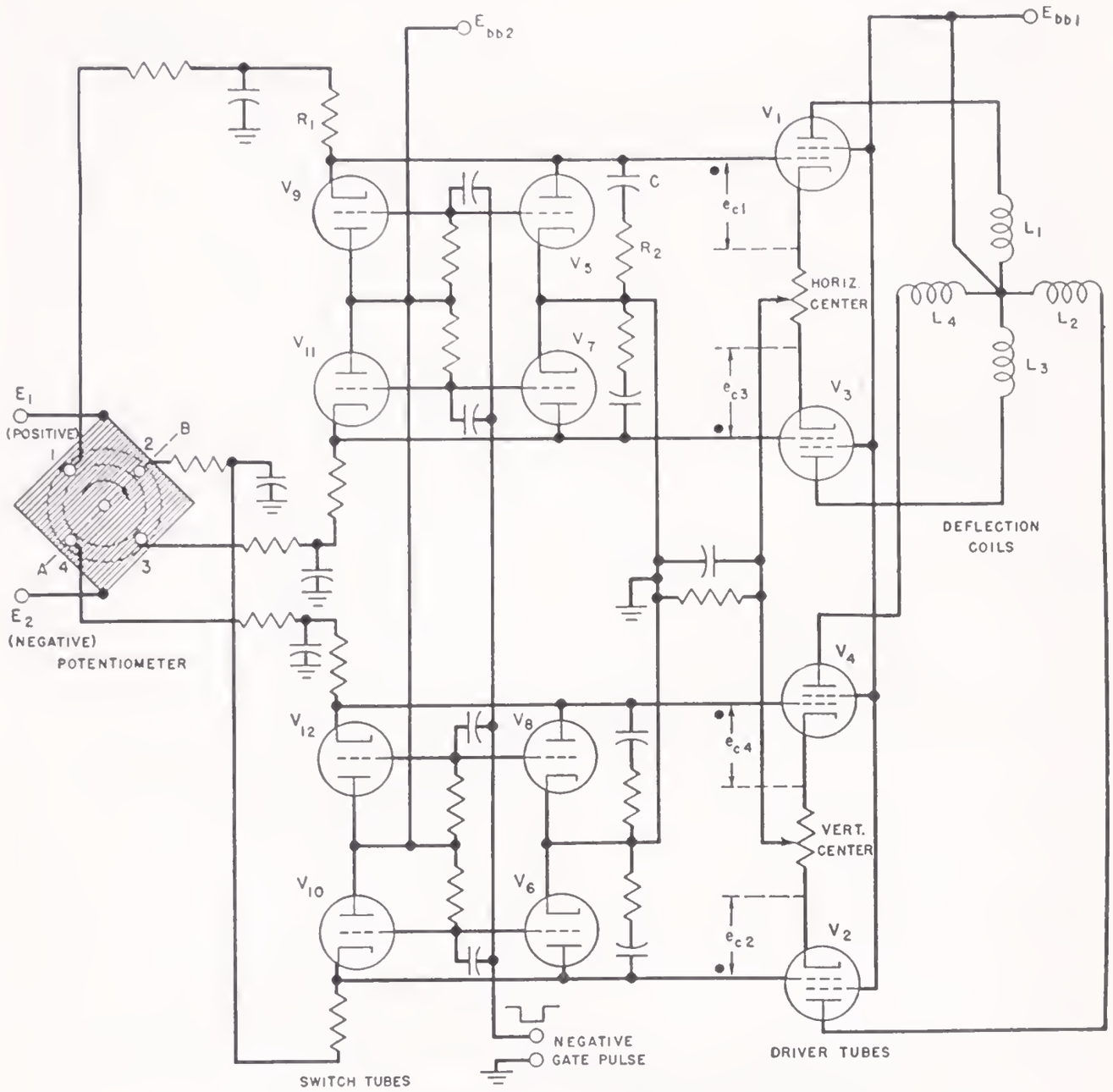


FIG. 41. Sweep circuit for plan-position indicator.

In the circuit of Fig. 41, each grid-voltage wave is supplied by a separate trapezoidal-voltage generator. For example, voltage e_{c1} is produced by the circuit comprising resistors R_1 and R_2, capacitor C, and switch tubes V_5 and V_9. This circuit differs from the trapezoidal-voltage generator of Art. 15 in two respects: (1) The supply voltage is obtained from brush 1 on the potentiometer instead of from a fixed source. As the potentiometer rotates, the supply voltage changes in magnitude and assumes negative as well as positive values. The variations of supply

voltage produce the sinusoidal variations of trapezoidal-wave amplitudes indicated in Fig. 42. (2) Two switch tubes are needed because of the reversal of polarity of the supply voltage. Between sweeps, tube V_5 conducts if the supply voltage is positive, and V_9 if this voltage is negative,

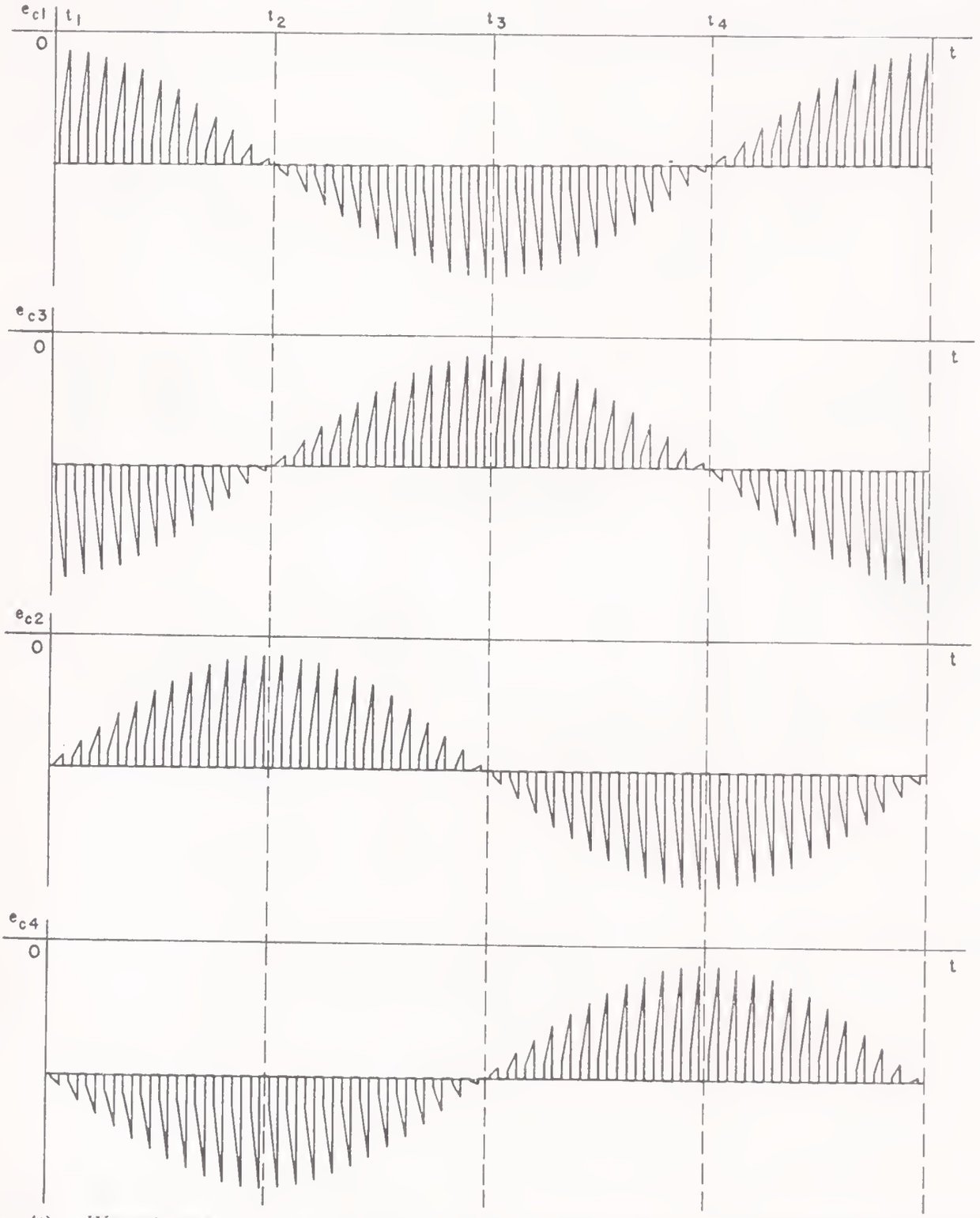


FIG. 42. Waveforms of grid voltages of driver tubes in Fig. 41. These waveforms apply when the antenna system rotates at constant speed.

thus preventing C from charging in either direction. During the sweep period, the negative gate pulse causes both tubes to be cut off, and the trapezoidal wave is formed.

Operation of the switch tubes may be explained more fully in terms of Fig. 43, which shows a portion of the circuit of Fig. 41. Between sweeps,

the grid voltage of V_5 is zero because of the positive grid return, but the grid voltage of V_9 is appreciably negative because of the plate voltage of V_5. The two switch tubes therefore may be considered to be unequal resistors forming a voltage divider across the E_{bb2} supply. Voltage E_{bb2} is small—100 volts or less—and e_{cn1}, the grid-to-ground voltage for V_1, is maintained at a fraction of this value, about 10 volts, by the conducting switch tubes. Because resistor R_1 is large, this between-sweeps value of e_{cn1} is essentially unaffected by changes of the voltage at the potentiometer brush.

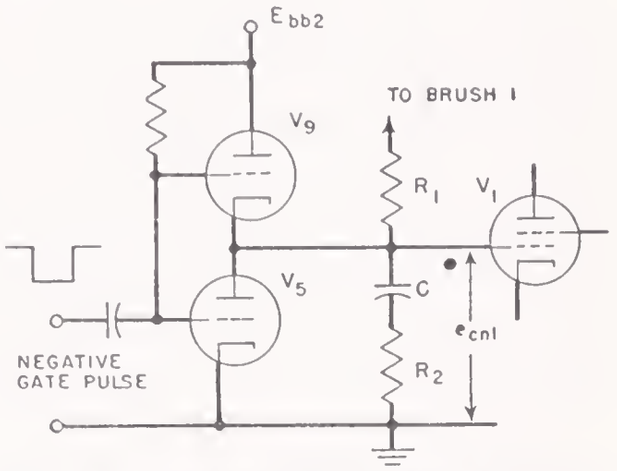


FIG. 43. Trapezoidal-voltage generator used with driver tube V_1 in Fig. 41.

Both switch tubes are cut off when the negative gate pulse is applied. A jump voltage appears because of R_2, and capacitor C begins to charge toward the voltage of brush 1. If this voltage exceeds the between-sweeps value of e_{cn1}, a positive-going trapezoid is produced; if the brush voltage is negative relative to the between-sweeps value, an inverted trapezoid is formed. When the brush voltage equals the between-sweeps voltage, as it does at instants t_2 and t_4 in Fig. 42, no trapezoid is produced.

Both the jumps and the slopes of the trapezoids produced are very nearly linear functions of the voltage obtained from the potentiometer. In order that the waves of Fig. 42 may be produced, therefore, it is necessary that the potentiometer supply four voltages which are sinusoidal functions of antenna angle and which have the same amplitude and phase relations as the envelopes of the waves in Fig. 42.

A special potentiometer is used to produce the required voltages. High-resistance wire is wound uniformly on a square card, and positive and negative voltages E_1 and E_2 are applied to the ends of the winding, as indicated in Fig. 41. The magnitudes of E_1 and E_2 are so chosen that the potential on the center line AB is the between-sweeps value of e_{cn1}. Potentials at points on one side of AB are positive and on the other side negative, relative to the between-sweeps voltage. The magnitude of the voltage difference between any point and line AB is proportional to the distance of the point from the line.

The construction of the potentiometer permits rotation of the card around its center point, connections to the end turns being made through slip rings. The potentiometer is arranged to turn with the antenna. As it rotates, the brushes follow a circular path on the card, and the distance of each brush from line AB varies sinusoidally. Thus a set of sym-

metrical four-phase voltages is obtained from the brushes, and these are the supply voltages required by the trapezoidal-voltage generators. The *R-C* filters shown in the leads from the potentiometer brushes in Fig. 41 are present because otherwise the small irregularities of voltage produced when a brush moves from one turn of wire to the next would produce irregularities in the trapezoidal waveforms of grid voltage.

The circuit of Fig. 41 is only one of several plan-position-indicator circuits used with fixed deflection yokes. Most commonly, these circuits employ a single trapezoidal-voltage generator together with a synchro or rotary transformer. This use of synchros is discussed in Art. 12 of Chap. V. The synchro may split the single trapezoidal voltage into four waves suitable as grid voltages for four driver tubes. Alternately, the synchro may follow the driver tubes and split a single saw-tooth current wave into two waves having the proper amplitudes for application to the horizontal- and vertical-deflection coils.

Sometimes electrostatic cathode-ray tubes with long-persistence screens are used in plan-position indicators. The presentation is then produced in the same general way as with fixed deflection coils. Instead of trapezoids, however, four variable-amplitude saw-tooth waves are produced and applied through two push-pull amplifiers to the four deflection plates.

CHAPTER V

SYNCHROS AND SERVOMECHANISMS

An arbitrary motion imparted to one shaft must often be reproduced accurately by the motion of a second shaft, sometimes with torque amplification. A direct mechanical connection between the *input* or *control shaft* and the *output* or *load shaft* may not be feasible either because of their separation distance or because of the need for torque amplification. When torque amplification is not required, extensive use is made of electrical *data-transmission* or *remote-indication* systems. The devices most commonly used for remote indication are *synchros*, which resemble small electric motors and operate on single-phase alternating current. When torque amplification is required, use is made of electrical or hydraulic devices called *servomechanisms*. Servomechanisms commonly include synchro units when control from a remote point is required.

Radar sets use remote-control systems to turn the antennas, to drive dials which indicate the antenna position, and to rotate the deflection coil of plan-position indicators in correspondence with the antenna. Remote-control systems are needed with high-accuracy radar sets to transmit the data collected by the set to the device that uses the data. For example, with wartime fire-control radar, a group of remote-control mechanisms transmits range, azimuth, and elevation coordinates of a target from the radar set to a computer. Another group of mechanisms aims the guns in accord with indications of the computer. Automatic-tracking radar sets include special servomechanisms that cause the antenna to follow a given object accurately and maintain a gate pulse at the range of the object.

A. SYNCHROS

A synchro is a small a-c machine used for the transmission of angular-position data. Synchros are known also by various trade names such as Selsyn, Synchrotie, Autosyn, and Diehlsyn. Other devices, such as Magnesyns[1] and potentiometer-and-magnet devices, are sometimes used for remote indication of angular position, but the most frequently used devices are synchros.

1. Construction and Connections of Synchro Motors and Generators. A simple example of the use of synchros is the system of Fig. 1 for

[1] R. S. Childs, "Magnesyn Remote Indication," *Trans. A.I.E.E.*, **63** (September, 1944), 679–682.

turning the deflection coil of a Type P indicator in accordance with the rotation of the radar antenna. The antenna is connected by gears to the shaft of the *synchro generator*. Voltages from the generator are transmitted to the *synchro motor* and cause the motor to turn in correspondence with the generator motion. Thus the deflection coil geared to the synchro motor is made to follow the antenna motion, and the direction of the range trace on the indicator tube corresponds to the direction of radiation from the antenna. For simplicity, the gear ratios are assumed to be 1:1 in this example, though in actual systems the synchros are usually geared to turn faster than the antenna and coil in order to provide increased accuracy (see Art. 11).

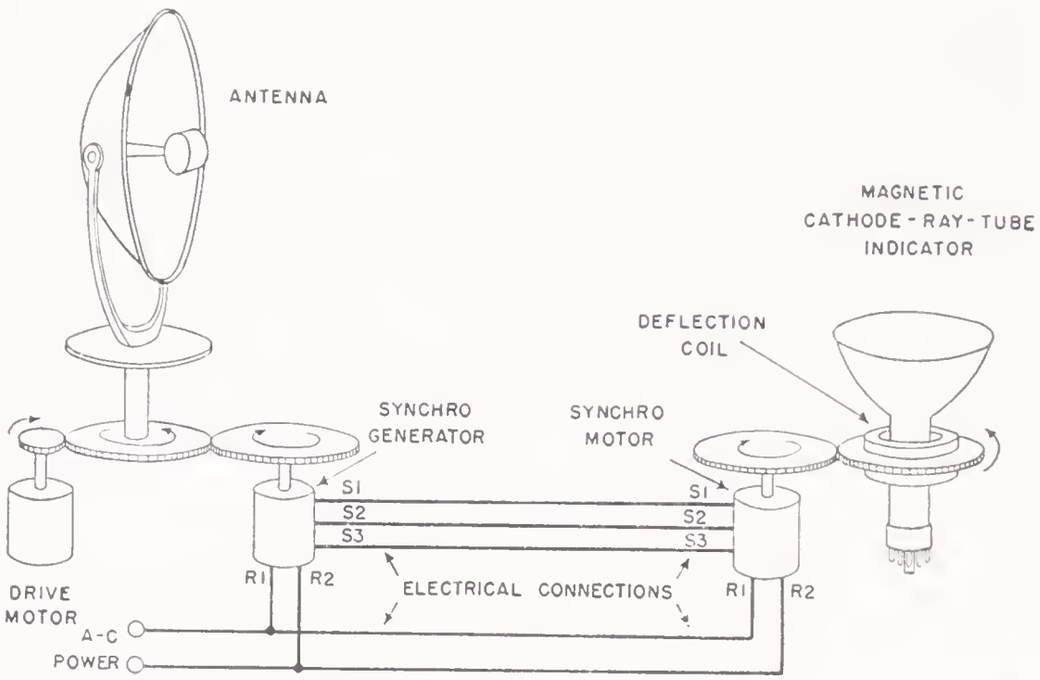


FIG. 1. Simple synchro system for turning deflection coil of Type P indicator.

The synchro generator is often called a *transmitter*, and the motor a *receiver* or *follower*. The generator and motor are identical in construction except that the motor is equipped with a mechanical oscillation damper (described in Art. 5) and with low-friction ball bearings whereas the generator does not have the damper and usually does not have ball bearings. As shown in Fig. 2, each machine has a stator (stationary part) and a rotor (rotating part) built of sheet-steel laminations and wound with coils of wire.

The rotor has two salient (projecting) poles and one winding. Connection to the winding is made through brushes resting on two slip rings, and the terminals are marked $R1$ and $R2$. The rotor shaft provides a mechanical input or output connection.

The stator is cylindrical and slotted and carries three similar windings, each distributed in several slots. In the synchro cross section of Fig. 3, one of the stator windings is shown occupying a group of four adjacent slots and another group of four slots located 180 deg from the first group.

The entire winding may be regarded as a coil, the turns of which lie in planes perpendicular to the line marked "Axis of coil A." When current flows in coil A, the magnetomotive force is in the direction of this axis.

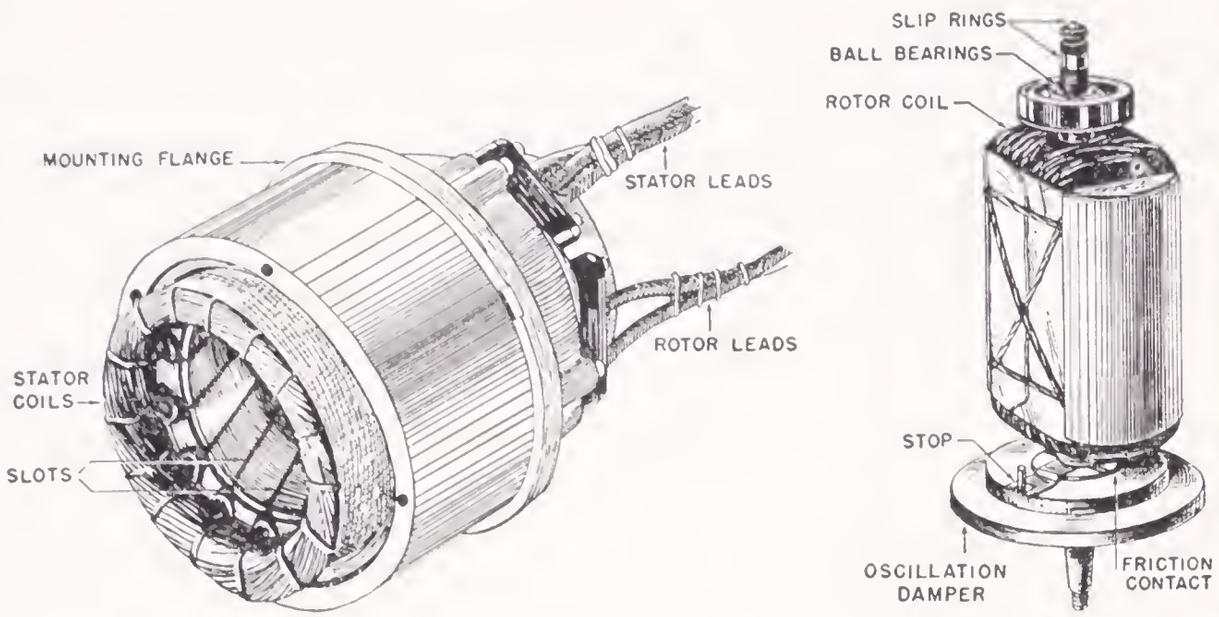


FIG. 2. Typical synchro.

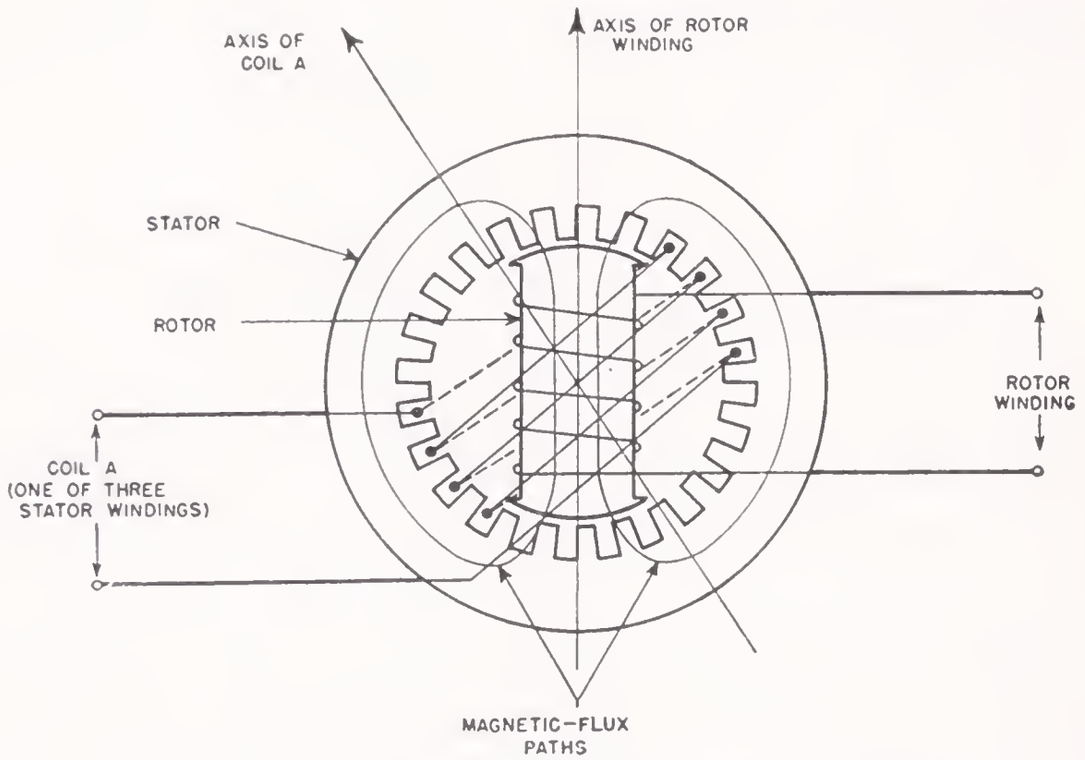


FIG. 3. Cross section of a synchro generator or motor.

Only one stator winding is shown in Fig. 3. Actually there are two others so placed that the axes of the three windings are 120 deg apart.

The three stator windings may be connected together in either wye (Fig. 4A) or delta (Fig. 4B). The three stator terminals are marked $S1$, $S2$, and $S3$, the numbers increasing in the counterclockwise direction around the stator. (Clockwise and counterclockwise apply to the

machine viewed from the shaft end and to circuit diagrams like Figs. 4*A* and *B*.) The circuit-diagram symbol for a synchro employing either the wye or delta connection is shown in Fig. 4*C*. The inner circle represents the rotor; the outer circle the stator. The arrows on the diagrams indicate the position of the rotor relative to the stator.

The electrical connections of the synchro system in Fig. 1 are as follows: Corresponding stator terminals of generator and motor are connected together, corresponding terminals of the rotor windings are connected together, and the rotor windings are connected to the a-c power line. These connections cause the motor to turn in the same direction as the

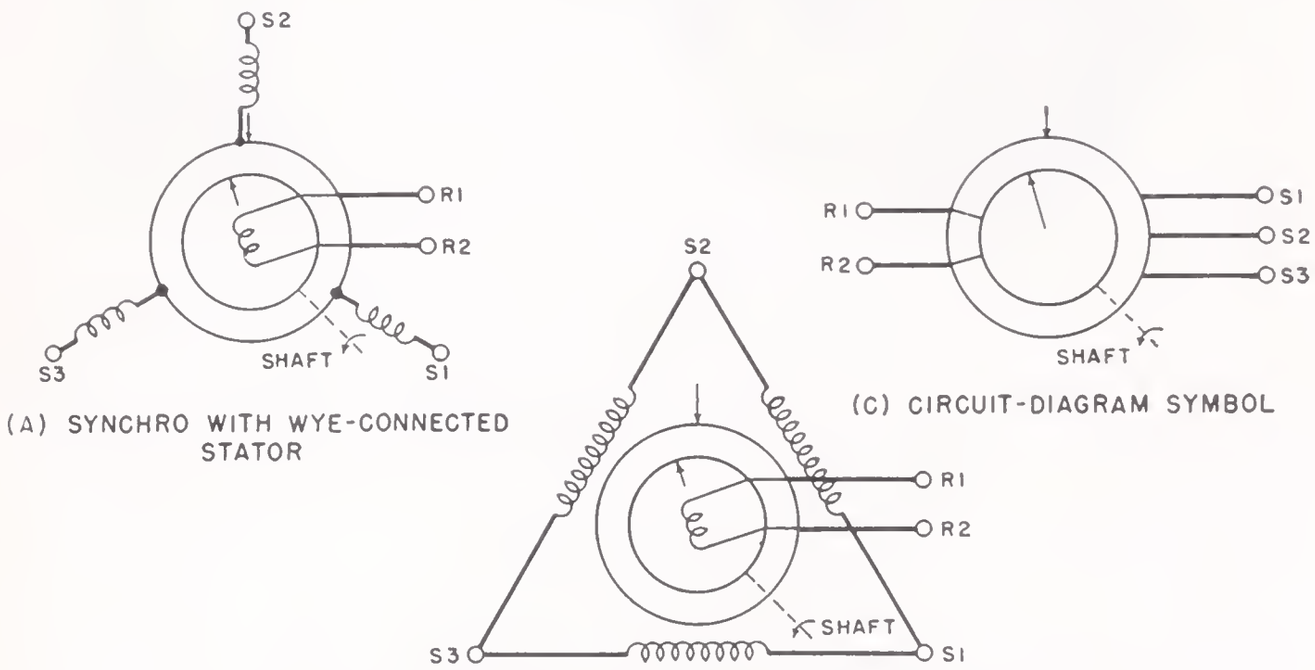


FIG. 4. Schematic representations of a synchro.

generator, as indicated by the arrows in Fig. 1. Altogether, five wires are usually run between generator and motor, though the number is sometimes reduced to four by using one wire in common for rotor and stator.

2. Operation of Generator—Electrical Zero. To understand why the motor follows the generator, the manner in which the generator stator voltages change as the generator shaft is turned must first be known. Suppose the rotor winding of the generator is connected to the a-c line[1] but the terminals of delta-connected stator windings are open, as shown in Fig. 5. A current flows in the rotor and causes a magnetic flux having a path somewhat as in Fig. 3. The flux varies sinusoidally with time at line frequency and induces alternating voltages in the stator windings.

[1] The power-supply voltage for synchro rotors is normally 115 volts, 60 cps for ground and shipboard installations, and 115 volts, 400 to 800 cps for aircraft installations.

The synchro is like a single-phase transformer in which the rotor winding is the primary and the stator windings are three secondary coils.

The magnitudes of the voltages induced in the stator windings depend upon the relative number of turns in rotor and stator coils and on the orientation of the rotor. For the position of the rotor in Fig. 5, voltage E_A across coil A has its maximum magnitude and is in phase with the line voltage E_{line}. As the rotor is turned clockwise from this position, E_A decreases and becomes zero when the rotor axis is perpendicular to the axis of coil A. Further turning of the rotor causes a voltage of reversed phase to appear across coil A.

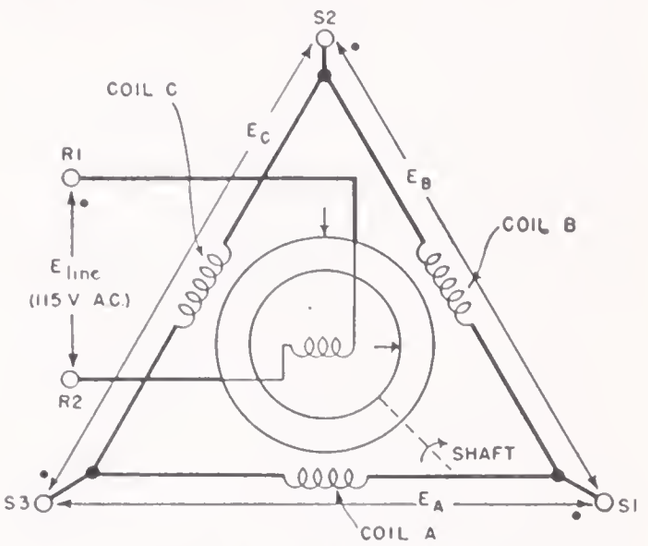


FIG. 5. Rotor and stator voltages of synchro with delta-connected stator.

The variation of the rms voltage E_A with the angular position of the rotor is plotted in Fig. 6.[1] The position of the rotor in Fig. 5 corresponds to $\theta = 90$ deg clockwise in Fig. 6. Positive values of E_A indicate an in-phase relation between E_A and E_{line}, and negative values a 180-deg

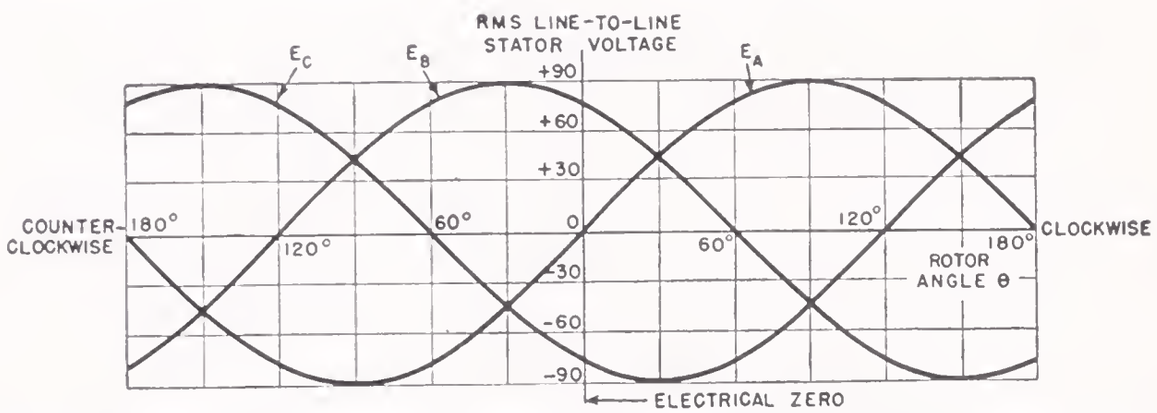


FIG. 6. Rms stator voltages as functions of rotor angle θ. Negative values indicate a phase reversal.

phase relation. The rms value of E_A varies sinusoidally with the rotor angle, because the distribution of the coil turns in the stator slots and the shape of the rotor pole faces are carefully chosen to secure such a variation. The curves of voltage across coils B and C are of the same sinusoidal shape. A displacement of 120 deg between the curves for any two coils appears because of the orientation of the coils. Point-by-point addition of the ordinates of the three curves in Fig. 6 shows that the

[1] The maximum stator voltage is shown in Fig. 6 as 90 volts, which is usual for Navy synchros. For Army synchros, the usual value is 105 volts, and for industrial synchros 55 volts is common.

sum of E_A, E_B, and E_C is zero for each position of the rotor; therefore, the delta may be closed as in Fig. 5 without appreciable circulating currents appearing in it.

The stator voltages do *not* constitute a three-phase set of voltages. Three-phase voltages are equal in magnitude and 120 deg apart in phase. The synchro stator voltages, on the other hand, are unequal in magnitude and are in phase or 180 deg out of phase with each other. If the rotor is turned slowly, the stator voltages are modulated with sinusoidal envelopes that differ in phase by 120 deg, as shown in Fig. 7. The individual cycles

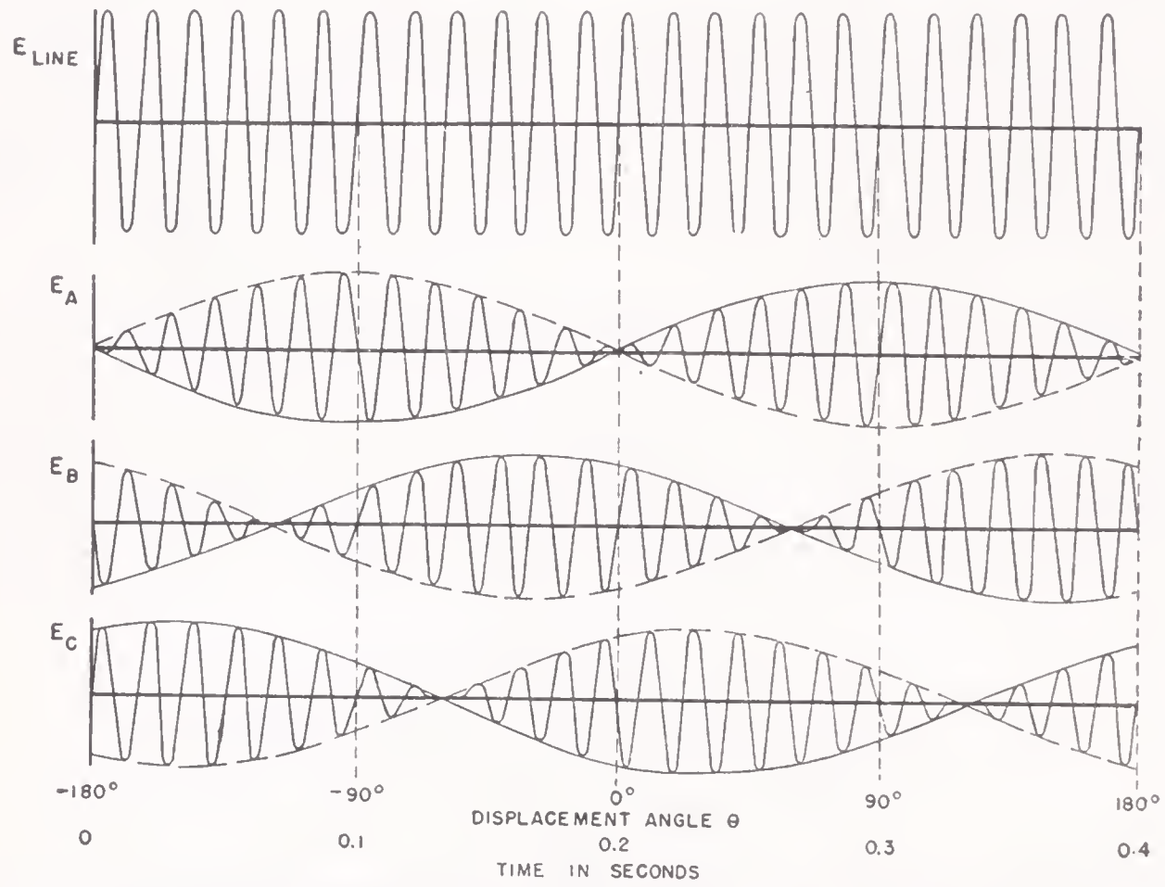


FIG. 7. Instantaneous rotor and stator voltages obtained as the rotor, connected to 60-cps line, is turned at 150 rpm.

of the stator voltages, however, are either in phase or 180 deg out of phase with the rotor voltage.

The angular position of the rotor is usually measured with respect to a standard position called *electrical zero*. By definition, electrical zero is the position of the rotor for which the voltage between $S1$ and $S3$ is zero, and the voltage at $S2$ (with respect to $S1$ or $S3$) is in phase with that of $R1$ (with respect to $R2$). The reference arrows on synchro diagrams indicate the zero position—one arrow is placed along the rotor axis, directed from $R2$ toward $R1$, and the other is so located on the stator that alignment of the two arrows indicates the electrical-zero position. The rotor angle θ is always measured from electrical zero. Note the electrical-zero position in Fig. 5, and note that the rotor angle θ is used in plotting the curves in Fig. 6.

To determine the electrical-zero position, connections are made as in Fig. 8. Electrical zero is the position in which E_A is zero and E_x is less than the line voltage. At a position 180 deg from electrical zero, E_A is again zero but E_x is more than 115 volts. A simpler method may be used if an electrical-zero autotransformer is available (see Art. 10). The rotor of the synchro and the autotransformer are connected in parallel to the power line, and the stator of the synchro is connected to the output terminals of the autotransformer; the synchro, if free to turn, then assumes its zero position.

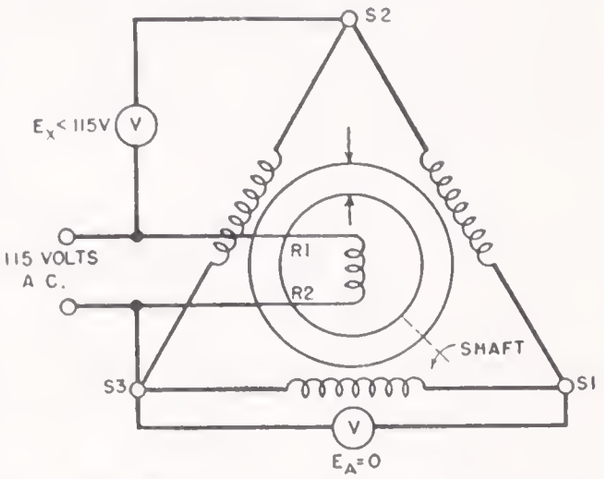


FIG. 8. Connections for determination of electrical-zero position.

The curves of Fig. 6 apply equally well to the wye-connected stator of the synchro in Fig. 9. To understand this, consider first the voltages E_1, E_2, and E_3 produced in the individual coils and plotted as a function of rotor angle in Fig. 10.[1] The voltages between stator terminals are the

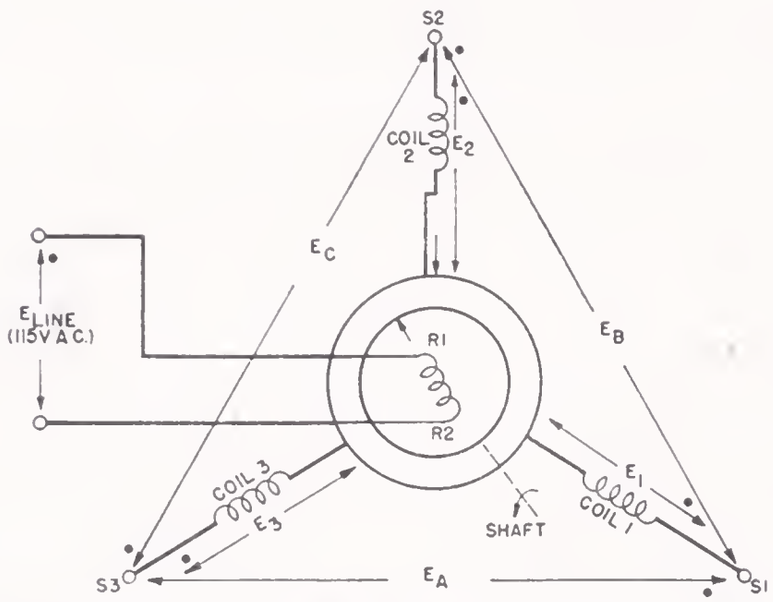


FIG. 9. Rotor and stator voltages of synchro with wye-connected stator.

differences of pairs of coil voltages; for example, E_A is $E_1 - E_3$. Point-by-point subtraction of ordinates of the curves of Fig. 10 shows that E_A, E_B, and E_C for the wye-connected machine vary with rotor angle in exactly the manner specified by Fig. 6. Fewer turns are placed on each coil of a wye-connected machine; thus the maximum values of the stator *terminal* voltages are made equal for the two connections.

[1] The maximum value of 52 volts applies to Navy synchros in which the maximum stator *terminal* voltage is 90 volts.

Because the two connections yield exactly the same stator voltages for each position of the rotor, it is difficult to tell which connection is used in a given machine. Delta-connected machines are assumed in the following discussions, but the results apply equally well for the wye connection. Both connections are in use.

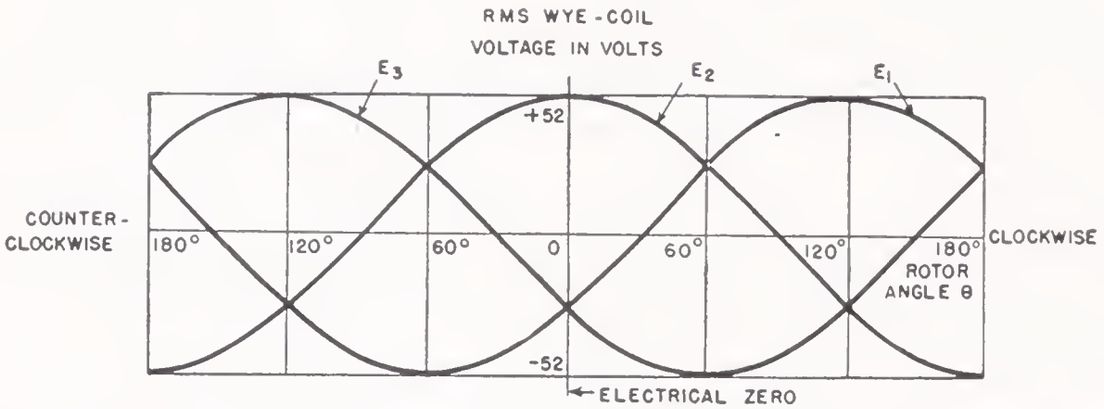


FIG. 10. Rms coil voltages as functions of rotor angle θ. Negative values indicate a phase reversal.

3. Operation of Motor. Consider, now, that the stator terminals of the generator are connected to corresponding terminals of the motor and that power is supplied to the generator rotor, as in Fig. 11. Assume temporarily that the rotor of the motor remains open-circuited. The voltages induced in the generator stator coils and applied to the motor

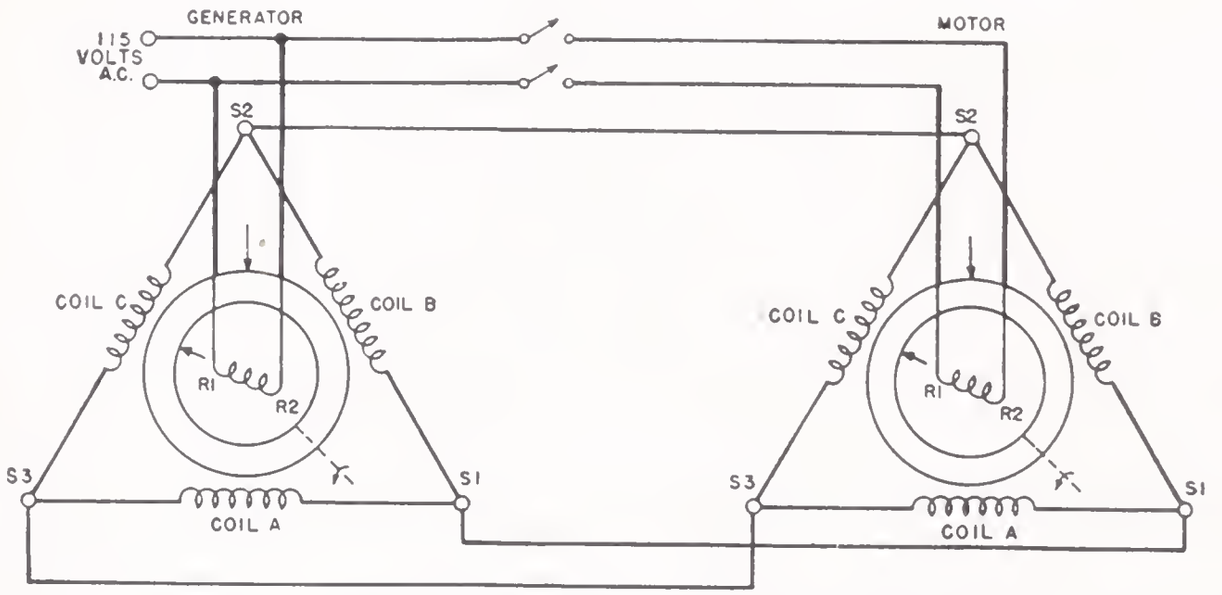


FIG. 11. Circuit for explaining operation of synchro motor.

cause currents to flow in the motor windings, and these currents set up magnetic flux in the motor. Connecting together corresponding stator terminals makes the self-inductance voltage induced by flux in each motor coil equal the voltage induced by flux in the corresponding generator coil, except for small voltage drops caused by the exciting currents flowing through the coil resistances. Since the motor and generator coils are alike, equal fluxes in corresponding coils are required to produce

equal voltages, and therefore the orientation of the magnetic field in the motor must be identical with that of the field in the generator.

The set of three stator voltages transmitted from generator to motor may be considered as an electrical signal that carries information about the position of the rotor in the generator. In fact, these voltages cause a magnetic field in the motor in a direction parallel to the magnetic field set up by the generator rotor.

A bar of soft iron placed in a magnetic field tends to align itself parallel to the field. Thus the salient-pole rotor of the synchro motor tends to turn into alignment with the field even though its winding is open. Operation with open rotor winding is undesirable, however, for three reasons: (1) For a given position of the generator, there are two stable positions of the motor rotor which are 180 deg apart. (2) The machines may become overheated because of excessive current. (3) The torque developed by the motor is small. These difficulties are avoided if the rotor winding of the motor is connected to the a-c line by closing the switch in Fig. 11. The rotor of the motor is now an electromagnet excited by alternating current and is said to be polarized. It has only one stable position. For the symmetrical connection, the stable rotor position of the motor corresponds exactly to that of the generator. If the rotor is connected to the a-c line and aligned with the rotor of the generator, the current circulating between the two synchro stators is zero, because the stator voltages of generator and motor oppose each other. The rotor currents are only the small currents needed to magnetize the machine.

Inasmuch as the synchro generator and motor are similar machines, the same reasoning could be followed in the reverse direction to show that the generator tends to align itself with the motor. Any torques developed in the two synchros at standstill are equal in magnitude and opposite in direction for the two machines. Each synchro tends to turn to the position of the other. In use, the generator is firmly driven, whereas the motor can turn freely and thus follows the generator motion.

The input to a synchro generator is the mechanical position of its shaft; the output is an electrical signal in the form of the three single-phase stator voltages. The synchro motor receives, as an electrical input, the three stator voltages and has a mechanical output, which is an angular rotation of its shaft.

The effect of changing synchro connections is illustrated in Fig. 12. For the symmetrical connection (Figs. 12*A* or 11), the motor assumes very nearly the same rotor angle as the generator and follows the generator with the same direction of rotation. A constant angular displacement of 180 deg may be obtained by a reversal of the connections to one rotor winding as in Fig. 12*B*, and a 120-deg displacement by a cyclic shift of stator connections as in Fig. 12*C*. Opposite directions of rotation are

obtained by interchanging stator connections. In Fig. 12*D*, a reversal is obtained by interchange of the connections to *S*1 and *S*3. In this way the motor and generator are made to reach the position of electrical zero at the same time. (Any other interchange of connections causes the motor to lead or lag by 120 deg.)

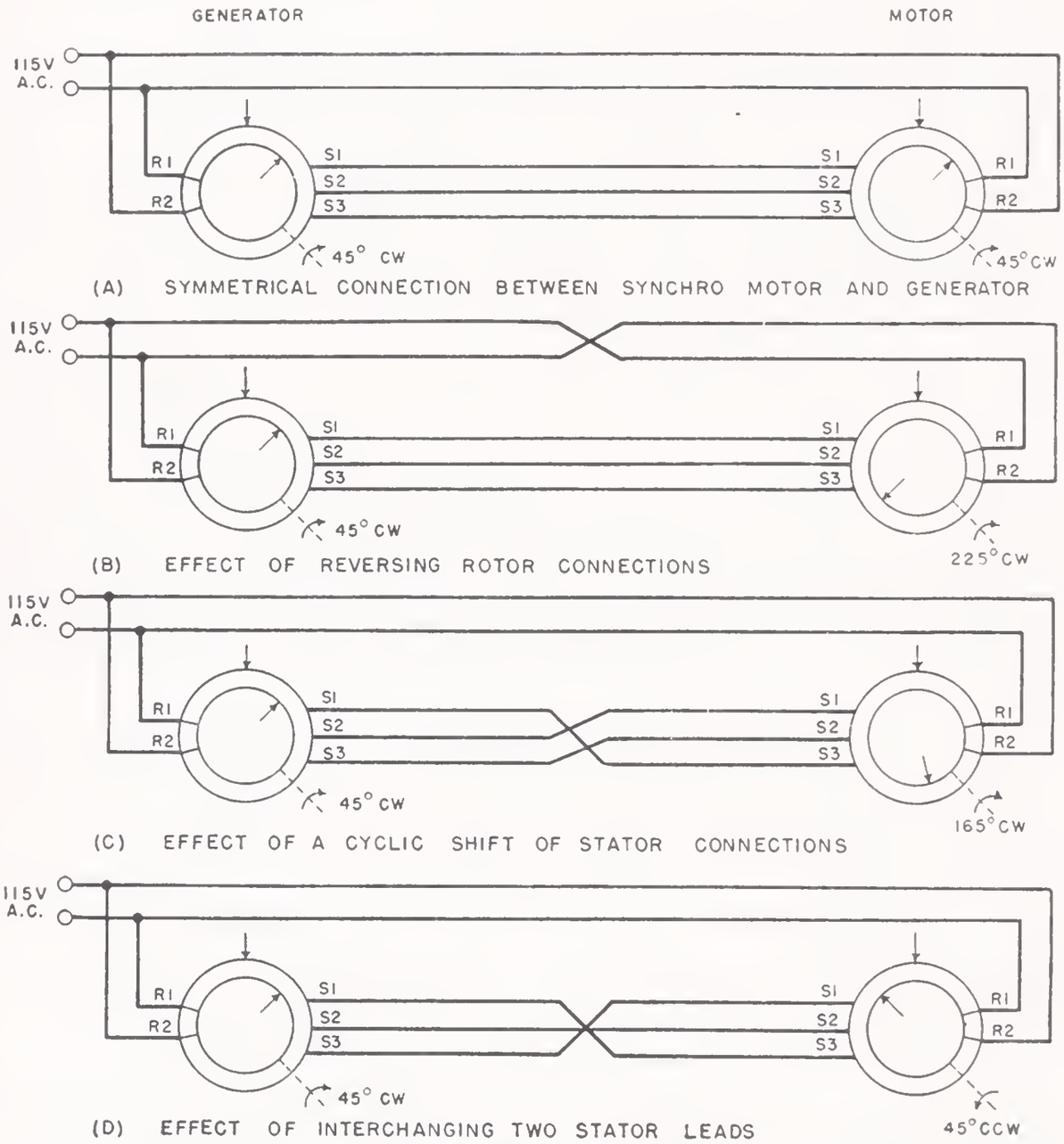


FIG. 12. Effect of changing connections between synchros.

One generator can be used to transmit position indications to two or more motors. For example, a gyrocompass aboard ship can be connected to a generator which is wired to a number of motors (compass repeaters) located at various points. If there are more than two motors, the generator should be larger than the motors. If any motor becomes jammed, it impairs the accuracy of all the other motors, since it becomes a generator competing with the real generator. Figure 13 shows a generator connected to three motors.

4. Torque-Angle Curve. In many applications of synchros, the accuracy of coincidence between generator and motor shafts is very important. Since both the synchro motor itself and the load connected to it have some friction and inertia (thus requiring torque during motion

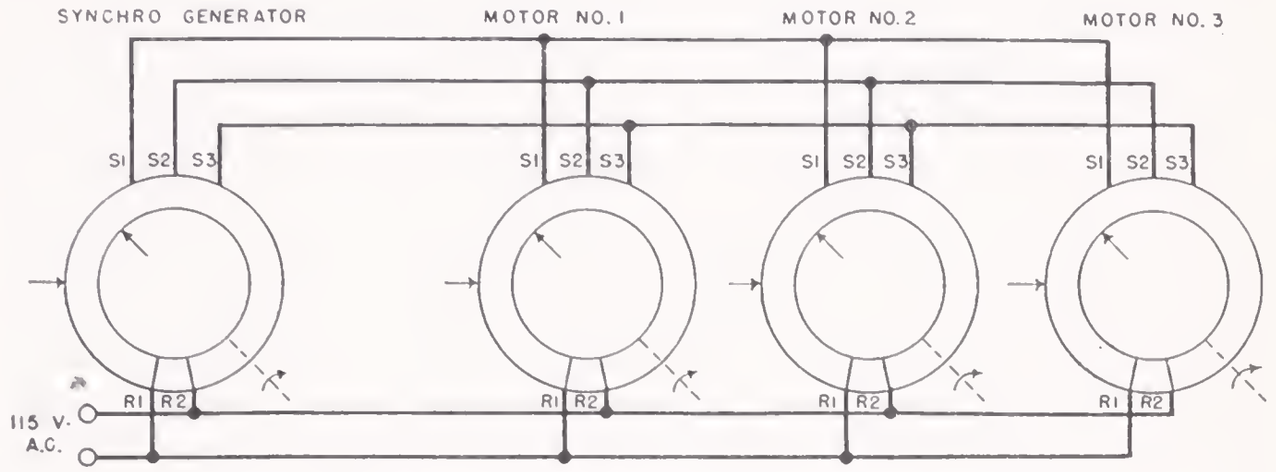


FIG. 13. One synchro generator controlling three motors.

and during acceleration), and since no torque is developed when the two rotors coincide, perfect accuracy is impossible.

The torque developed for any given error or displacement angle δ may be determined from a torque-angle curve such as that of Fig. 14. This curve is very nearly sinusoidal, so that the point of maximum torque occurs at an angle of approximately 90 deg. As indicated in the figure, the direction of the torque depends upon the sign of the displacement angle. The curve applies for standstill conditions, and under these conditions motor and generator torques are equal. At high speeds of rotation, the curve changes somewhat and motor and generator torques become unequal. A curve of rotor current versus displacement angle is also given in Fig. 14.

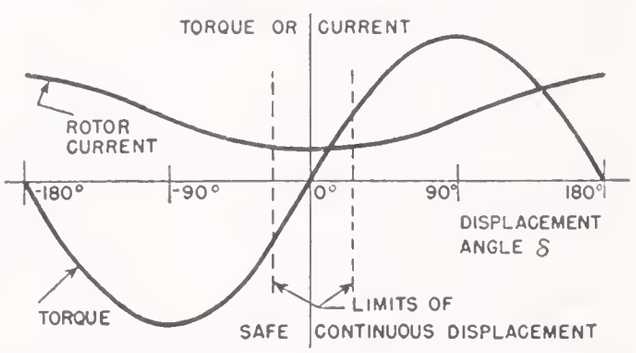


FIG. 14. Torque and rotor current as functions of displacement angle.

If the load on the motor should require a torque greater than the maximum or pull-out torque, the motor would drop out of step with the generator. Normally the loads are so small that the displacement does not exceed 10 or 20 deg. Operation with light loads is desirable (1) for accuracy of angle transmission, (2) so that momentary overloads of several times normal loads will not pull the motor out of step, and (3) to prevent overheating of the machines. Although machines could be designed so as not to overheat when operated continuously at large displacement angles, there is no reason for doing so because accuracy requires

the normal displacement to be small. The maximum safe continuous displacement is 20 deg in the usual designs.

If the resistance of the wires connecting the two stators is high, the ordinates of the torque-angle curve are reduced. The resistance of the wires connecting the rotors has a similar but smaller effect, because the rotor voltage is usually higher than the stator voltage.

Except for the small errors caused by friction, especially static friction, in the ball bearings of the motor, the no-load error of a synchro system would seem to be zero, according to the curves of Fig. 14. Actually, however, there are errors even at no load, because the two machines cannot be made exactly alike, and therefore the relative positions of the rotors vary a little as the rotors turn through a revolution. To make these errors small (for typical values see Tables 1 and 2, Art. 8), the machines are carefully designed and built. The motor rotor is accurately balanced. The inner surfaces of the stators are ground to true cylinders, and the rotors are accurately centered therein to minimize torque arising from magnetic dissymmetry. The number and size of the slots and the shape of the pole faces are carefully chosen to make the stator-voltage curves vary sinusoidally with rotor angle. To minimize any tendency of the rotor to lock in at one position due to tooth torque and to ensure smooth operation, either the pole tips or the slots are skewed. Synchros of high accuracy require much more careful workmanship than a fractional-horsepower motor of the same size, and their cost is correspondingly greater.

The mechanical system of Fig. 15 has nearly the same torque-angle curve as a synchro system and therefore serves as an analogue for use in explanations of synchro operation. Two shafts, representing the generator and motor shafts, are coupled by means of a spring joining crank arms on each shaft. The angular displacement δ between the two cranks increases as the transmitted torque increases. The torque is zero at zero displacement and again at 180-deg displacement and greatest near 90 deg. Beyond 90 deg the spring tension increases but the component of tension perpendicular to the crank arms decreases; hence the torque decreases. The spring tension is analogous to the stator currents of the synchro system. The 180-deg position is obviously an unstable one.

5. Oscillation Dampers. If the generator is given a sudden angular displacement, the motor—because of its inertia—tends to overtravel and oscillate about its final position. This tendency should be clear from the mechanical analogue of Fig. 15. In order to minimize the duration of oscillations, as well as to overcome any tendency of the motor to start *spinning*—that is, running at high speed as an induction motor—the motor is usually equipped with a mechanical oscillation damper.

The most obvious method of damping would be to use bearings with

high friction. This method of damping would be disadvantageous, however, because in order to rotate, even though at constant speed, the motor would have to develop a torque to overcome the bearing friction, and an error angle would result. Damping is obtained without introducing such an error by using a so-called inertia damper. In one form the inertia damper consists of a flywheel having a loose rim. When the acceleration is high—as it is during an oscillation—the rim slips on the central part and the friction between the two parts produces a damping torque. When the velocity is constant, however, the rim turns with the central part and there is no friction or error due to friction. An equivalent inertia damper is a hollow flywheel filled with mercury.

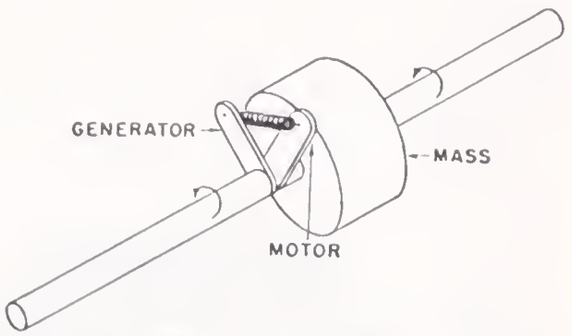


FIG. 15. Mechanical analogue of synchro system.

In another form of inertia damper, a flywheel is mounted on the shaft (see Fig. 2*B*) so that the flywheel turns freely on an intermediate ring or bushing through a small angle, after which it strikes a stop on the bushing. Any further rotation of the flywheel relative to the shaft causes the bushing and flywheel to turn together. Since there is a friction disk between the bushing and the shaft, this additional motion introduces appreciable damping. This arrangement permits oscillations of small amplitude to occur without much damping, but oscillations of larger amplitude are damped rapidly.

6. Synchro Control Transformer. Not all synchros are used in the manner of a generator and motor to repeat a shaft motion. The *synchro control transformer* (sometimes called a *synchronous transformer*) is used with a synchro generator to indicate the difference between the angular positions of the two shafts, a use that has important applications in electrical servomechanisms. The control transformer does not exert a torque to tend to align itself with the generator. Instead it produces a voltage, sometimes called an *error signal* or *error voltage*, which indicates the error of alignment between its shaft and the generator shaft.

The synchro control transformer has a stator similar to that of a synchro generator or motor, except that the three stator windings of the control transformer have more turns of finer wire. Its rotor, shown in Fig. 16*A*, is cylindrical rather than salient-pole and has a single high-impedance winding consisting of a large number of turns of very fine wire. The rotor winding is placed in slots as shown in the figure. Stator and rotor terminal designations and the positions of reference arrows are indicated in Fig. 16*B*. Note that the rotor reference arrow is placed at right angles to the axis of the rotor winding rather than in the rotor

axis; otherwise the diagram is the same as for a synchro generator or motor.

The control transformer operates as a single-phase transformer having three stationary primary windings and one movable secondary winding. The circuit diagram of a system using a generator and a control transformer is shown in Fig. 17. The stator voltages applied from the synchro

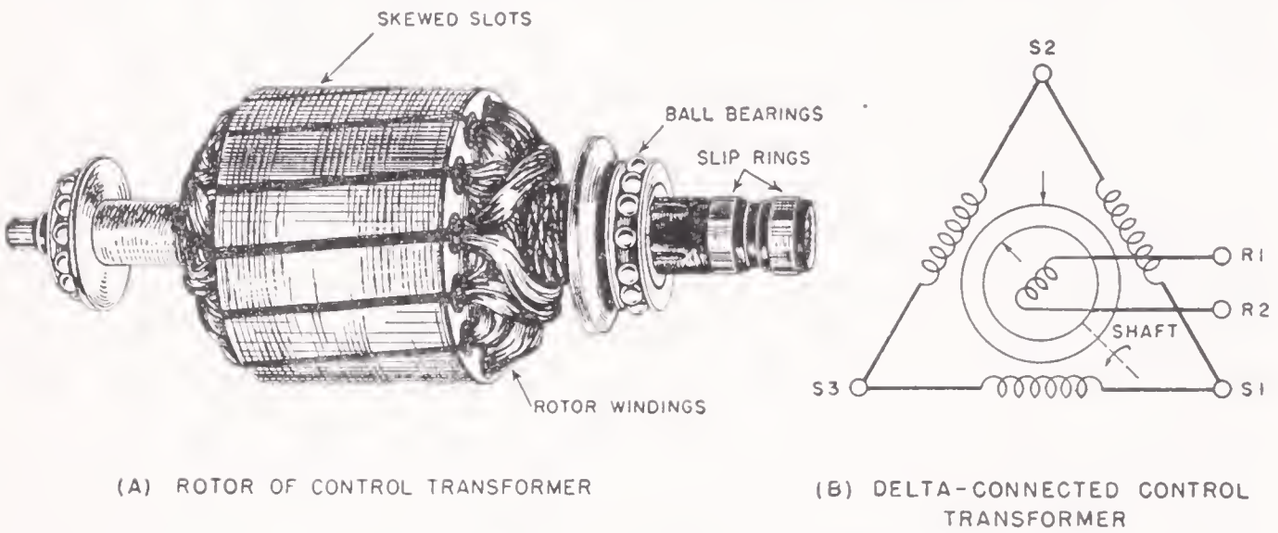


FIG. 16. Synchro control transformer.

generator to the control transformer establish a magnetic flux Φ in the control transformer. If the generator rotor angle is θ_i, then the flux is displaced from the stator reference by an angle θ_i in both machines. The rms value of the voltage induced in the rotor winding of the control transformer depends upon the rotor position relative to the flux—no voltage is produced if the coil axis is perpendicular to the flux, and a maximum voltage is obtained if the axes are parallel.

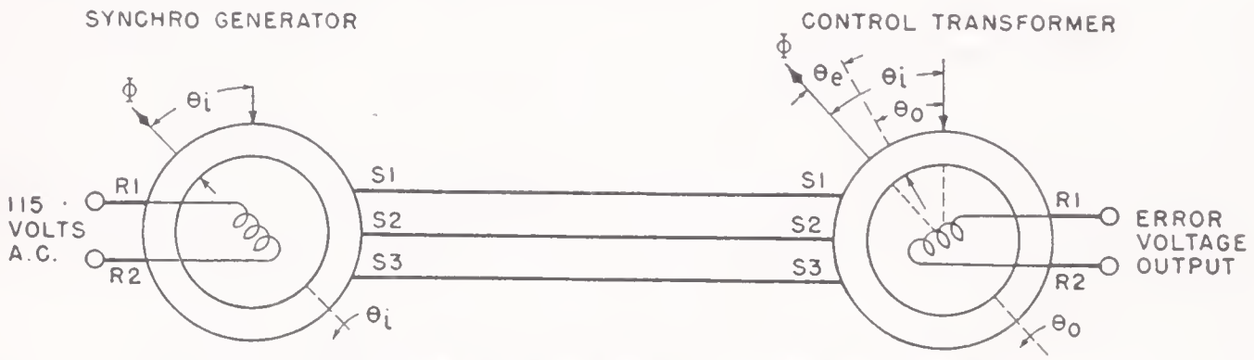


FIG. 17. Control transformer used to compare two shaft motions.

Electrical zero for a control transformer is one of the two zero-rotor-voltage positions obtained when the stator is connected to a generator that is set to its electrical zero. A counterclockwise rotation of the control transformer through less than 180 deg from the electrical-zero position causes a voltage at $R1$ (with respect to $R2$) in phase with the voltage at $S2$ (with respect to $S1$ or $S3$). This corresponds to a control-transformer rotor voltage in phase with the generator rotor voltage, the voltages being measured at corresponding terminals. Reference arrows

are placed on circuit diagrams of control transformers to indicate electrical zero by their coincidence. The rotor angle θ_o (see Fig. 17) is measured from the electrical zero position.

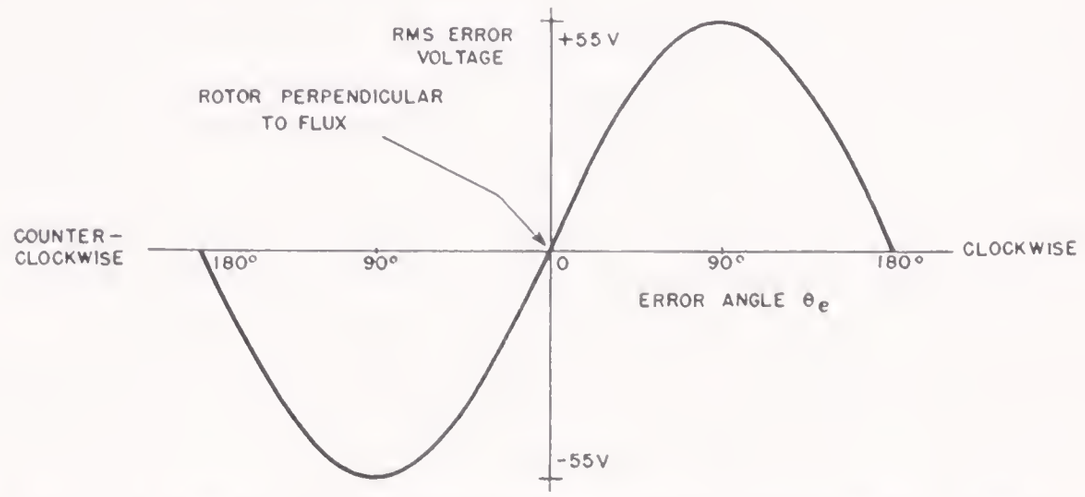


FIG. 18. Error voltage of control transformer as function of error angle. Negative values of voltage indicate reversed phase.

The value of the error voltage induced in the rotor varies with the error angle $\theta_e = \theta_i - \theta_o$, as shown in the curve of Fig. 18.[1] A positive sign for error voltage on this curve indicates that the error voltage and line voltage are in phase, and a negative sign indicates that the two are 180 deg out of phase. A very nearly sinusoidal error-voltage curve is obtained by appropriate distribution of coil turns among the slots.

The error voltage as a function of time is plotted in Fig. 19 for clockwise and counterclockwise error angles. Note that it is an alternating voltage of line frequency either in phase or 180 deg out of phase with the line voltage.

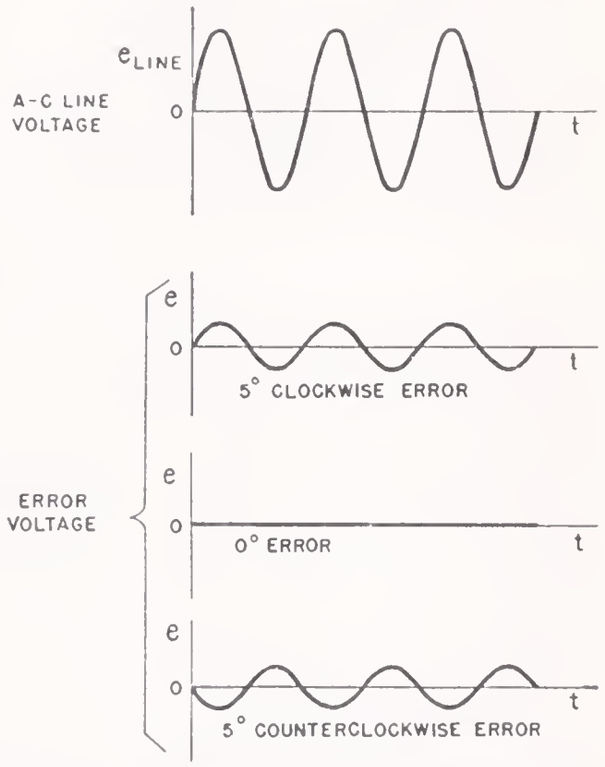


FIG. 19. Relation of error voltage to line voltage and error angle.

As commonly used in electrical servomechanisms, the synchro control transformer supplies an error signal from its rotor winding to an amplifier that controls a d-c or a-c motor. The circuit is such that the speed of the motor is proportional to the magnitude of the error voltage, and the direction of rotation is determined by the phase of the error voltage with respect to that of the line voltage. In the normal operation

[1] The maximum value of 55 volts applies to a standard Navy control transformer.

of such a system, the error angle is kept small, so that the error voltage is almost directly proportional to the error angle.

Since the amplifier circuit to which the control-transformer-rotor output is normally connected has a high input impedance, the rotor current is negligible. Because the rotor currents are so small, and because the rotor is cylindrical, the torque developed in the control transformer is negligible, no matter to what position the rotor is turned. The cylindrical rotor and the high-impedance stator windings serve to make the exciting currents in the stator windings small. (If a salient-pole rotor with its axis 90 deg from the field axis were used, the reluctance of the flux path would be high and large stator currents would flow.)

7. Differential Synchro. If it is desired to have the angular position of an output shaft equal the *sum* or *difference* of the angular positions of two input shafts, a *differential synchro* may be used.

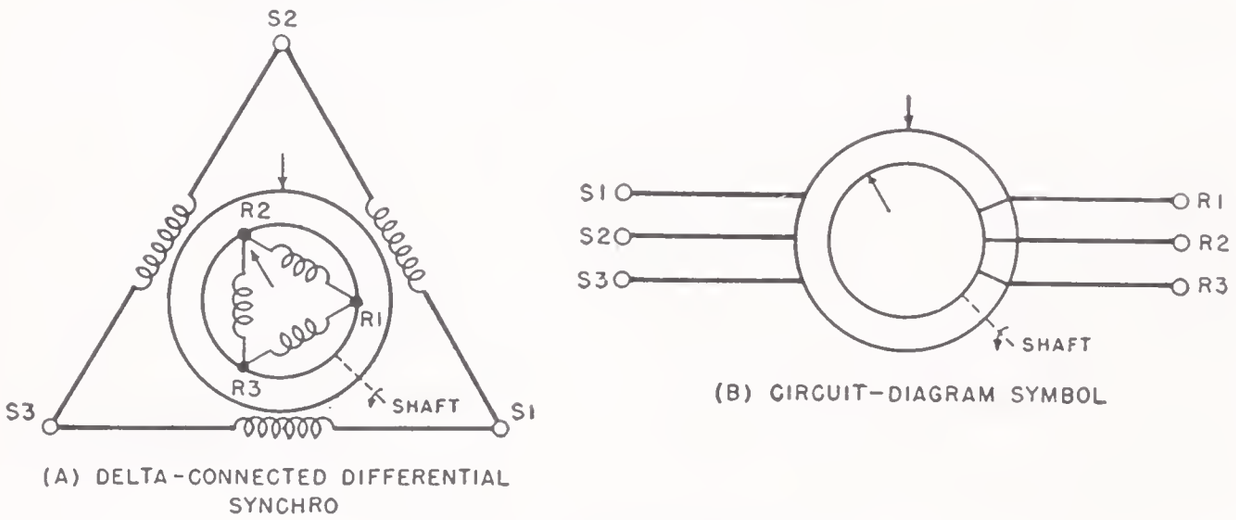


FIG. 20. Differential synchro.

The differential synchro has a stator identical with that of a synchro generator or motor. It differs from the generator or motor in that it has a cylindrical, instead of a salient-pole, rotor; and the rotor, like the stator, has three distributed windings spaced 120 deg apart. The turns ratio between stator and rotor windings is approximately 1:1. The rotor windings have a slightly larger number of turns than the stator windings, so that, with voltages applied to the stator, the extra turns on the rotor compensate for internal impedance drops and give a 1:1 ratio between rotor and stator terminal voltages. Three slip rings and brushes are used for connections to the rotor windings. Terminal designations are shown in the schematic representation of Fig. 20. Both wye and delta connections are used.

The electrical zero of a differential synchro is the position in which corresponding coil axes of rotor and stator are parallel. A stator reference arrow is placed on diagrams, such as Fig. 20, on the electrical-zero axis of the stator, that is, along the axis of the flux set up by the stator

windings when an electrical-zero signal is applied. Similarly, a rotor arrow indicates the electrical-zero axis of the rotor. The two arrows are aligned when the differential synchro is in its electrical-zero position.

There are two types of differential synchros: the *differential generator* and the *differential motor*. They have the same construction except that the differential motor has an inertia damper like that of an ordinary synchro motor. The differential generator is normally used as an intermediate unit connecting the output of a synchro generator to the input of a synchro motor or control transformer. Its purpose is to receive an electrical signal corresponding to a certain angular position of the generator, to modify this signal by an amount corresponding to the angular position of its own rotor, and to transmit the modified electrical signal to the output synchro. This modified electrical output signal produces an angular position of the flux in the output synchro which is either the sum or the difference of the rotor angles of the two generators. The differential motor has two electrical inputs from separate synchro-generator units, one to the stator and one to the rotor winding, and a mechanical output which is its shaft position.

Consider first the differential-generator system of Fig. 21*A*. The stator voltages of the generator are applied to the stator of the differential generator and establish a magnetic field in the differential having the same orientation as the field in the generator. In Fig. 21*A*, the field in the synchro generator is displaced 45 deg counterclockwise from the generator stator reference because of the rotor angle. Therefore the magnetic field in the differential is displaced 45 deg counterclockwise *from the stator reference*, as indicated by the flux arrow Φ.

The rotor voltages of the differential generator depend upon the position of the magnetic field relative to the rotor coils in exactly the same way that the stator voltages from an ordinary synchro generator do. In Fig. 21*A*, the rotor is turned 90 deg counterclockwise from electrical zero so that the flux arrow is 45 deg clockwise from the rotor reference arrow. Therefore the rotor voltages of the differential generator cause a magnetic field 45 deg clockwise from the stator reference of the synchro motor, and the rotor of this motor tends to align itself with this field.

The differential generator in this example modifies the original electrical signal for a 45-deg counterclockwise rotation by subtracting a 90-deg counterclockwise rotation, with the result that the motor assumes a position 45 deg clockwise from electrical zero. Thus, in general, this connection makes the rotor angle of the motor equal that of the generator minus that of the differential generator.

Different relative directions of rotation can be obtained by interchanging pairs of differential generator input or output leads as illustrated in Fig. 21. For example, in Fig. 21*B*, connections to terminals 1 and 3 are

reversed at the rotor and at the stator of the differential, and the rotor angle for the motor is the sum of the generator and differential angles. For the angles indicated in the figure, the flux in the generator is 45 deg counterclockwise from the stator reference arrow. Because of reversed connections, the differential generator flux is 45 deg clockwise from its

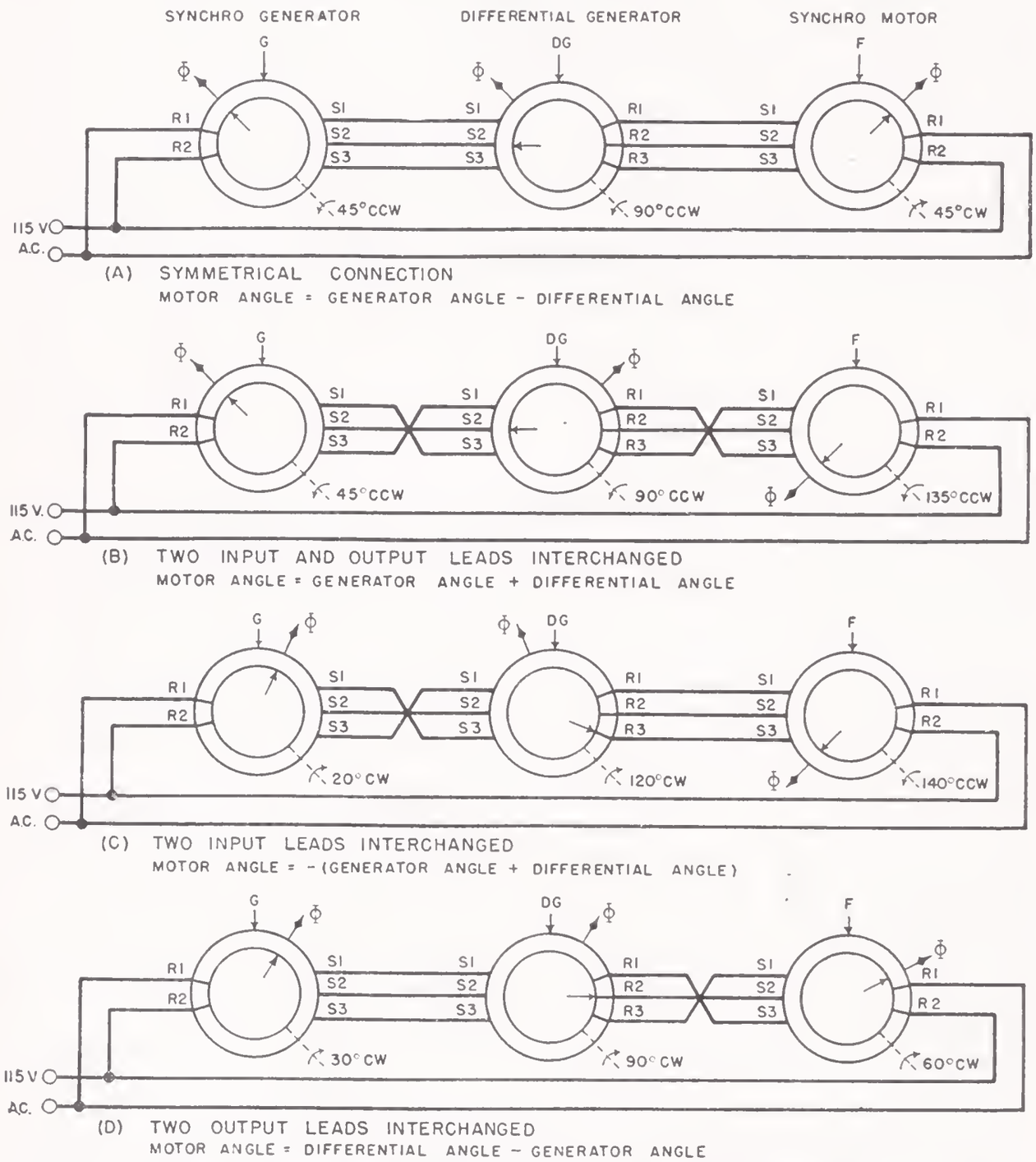


FIG. 21. Effect of interchanging input or output leads of differential generator.

stator reference or 135 deg clockwise from the rotor reference. The second reversal of connections makes the motor flux take a position 135 deg counterclockwise from the stator reference.

The rotor windings of the differential generator in Fig. 21 may, of course, be connected to a control transformer instead of a motor. The control-transformer output voltage is then a measure of the error between

its own rotor angle and the combination of generator and differential-generator rotor angles that positions the control-transformer flux.

Consider now the synchro system with a differential motor shown in Fig. 22. Voltages indicating a 45-deg rotation are applied from synchro generator A to the stator windings of the differential motor, and a magnetic field Φ_S is created along an axis 45 deg counterclockwise from the stator reference arrow. The 60-deg clockwise electrical signal from synchro generator B is applied to the rotor windings of the differential motor and establishes a magnetic field Φ_R that is 60 deg clockwise from the rotor reference arrow. The rotor is equivalent magnetically to a single-winding rotor having an axis 60 deg clockwise of the rotor reference arrow. Therefore the motor, if free to turn, assumes the position shown

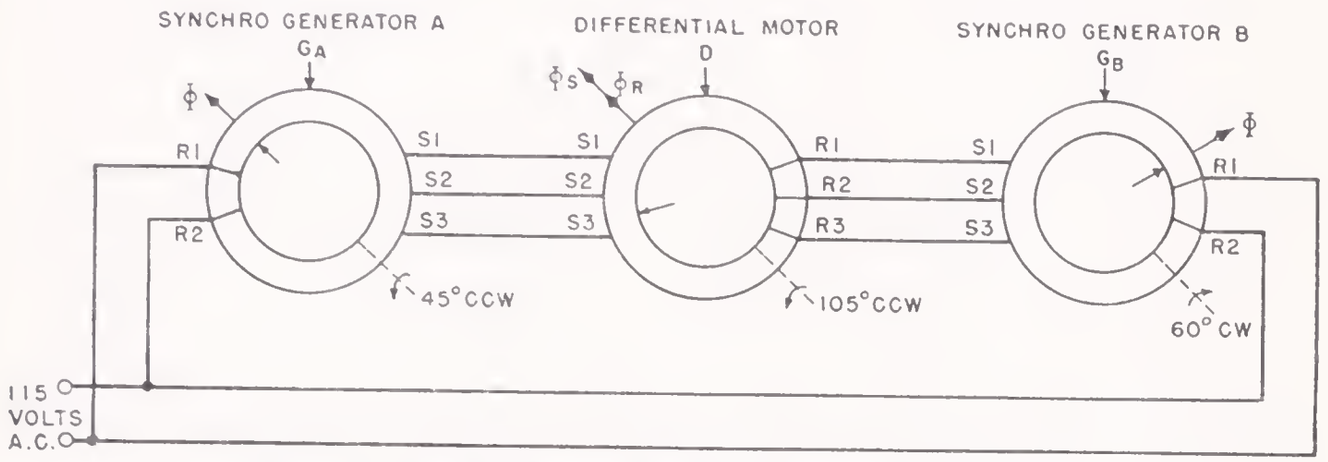


FIG. 22. Synchro system including a differential motor.

in Fig. 22 in which Φ_S and Φ_R are aligned. The rotor angle of the differential motor is 105 deg counterclockwise; that is, the motor angle is the angle of generator A minus the angle of generator B. Reversing pairs of connections at the differential can change the relative directions of motion in much the same way as illustrated in Fig. 21 for the differential generator.

8. Designations of Synchros. Synchros used in Army and Navy equipment are given standard designations. These synchros are frequently encountered in radar systems.

Navy synchros are designated by an Arabic numeral followed by one or two letters, for example, 5DG. The numeral indicates the size and the letters define the function of the synchro. The meaning of the letters is as follows:

G—Generator	CT—Control transformer
DG—Differential generator	S—Special unit
F—Motor (follower)	H—High-speed unit (speed greater than 300 rpm)
D—Differential motor	

The meaning of the size numbers is indicated by the approximate weights and dimensions in Table 1, where the maximum no-load error is also given. All sizes are built for operation at a line frequency of 60 cps, with rotor voltages of motors and generators of 115 volts, maximum stator voltages of 90 volts, and maximum rotor voltages of control transformers of 55 volts. Although Navy synchros are built by several manufacturers, standard dimensions and performance characteristics are adhered to by all. Sizes 1, 5, and 6 are most common in radar equipment.

TABLE 1. NAVY SYNCHROS

Size	*Approximate weight, lb*	*Approximate length, in.*	*Approximate diameter, in.*	*Maximum no-load error, deg*
1	2	3.9	2.2	1.5
3	3	5.2	3.1	
5	5	6.0–6.8	3.4–3.6	0.6
6	8	6.4–7.5	4.5	0.6
7	18	8.9–9.2	5.8	0.6
8	60	13.1	8.6	

Army synchros are designated by Roman numerals which denote both size and function. These synchros are designed for 50- and 60-cps service. A few typical types of Army synchros are listed in Table 2.

TABLE 2. ARMY SYNCHROS

Type	*Function*	*Approximate weight, lb*	*Maximum no-load error, deg*
I	Generator	11.8	0.5
II	Motor	10.3	0.5
IV	Generator	4.8	0.5
V	Motor	4.8	0.5
X	Motor	1.3	2.0
XIV	Differential motor	4.8	
XV	Control transformer	4.8	

9. Synchro Capacitors. When a synchro generator is connected to a differential generator or control transformer, it must supply an excitation current to the other machine, even if no error angle exists. The exciting current is the sum of two components: (1) a loss current, which is in phase with the voltage and which supplies the heat loss in the copper of the windings and in the iron, and (2) a magnetizing current, which lags

the voltage by 90 deg and establishes the magnetic field. The magnetizing component is normally the larger of the two.[1]

To obtain maximum accuracy from a synchro system, the currents drawn from the synchro generator should be kept as small as possible. A reduction of the current supplied to a differential generator or control transformer is accomplished by the use of *synchro capacitors*, sometimes called *capacitor exciters* or simply *exciters*. A synchro-capacitor unit comprises a delta-connected set of equal capacitors. These capacitors are connected in parallel with the stator windings of the differential generator or control transformer. The 90-deg leading current of the capacitor balances out the 90-deg lagging magnetizing component of the exciting current of the synchro. Under no-load conditions, therefore, the line current is reduced to the synchro loss current.

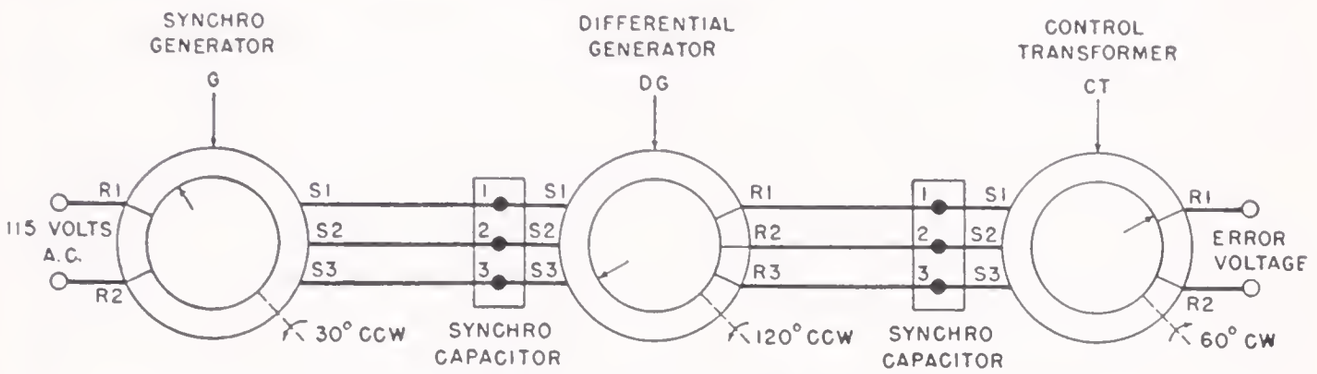


FIG. 23. Connection of synchro capacitors in a system.

A synchro capacitor is connected as close as possible to the terminals of the differential generator or control transformer to which it supplies magnetizing current. This is done to reduce the currents in the long leads to the synchro generator. Figure 23 shows the capacitor exciter connected in a typical synchro system.

Synchro capacitor units are made in a variety of sizes for use with different types and sizes of synchro units. They are usually rated according to the total capacitance, which is the sum of the three individual capacitances. Standard synchro capacitors should always be used with synchros, as they consist of very high-grade paper-foil units which are carefully matched so that the individual values differ by less than 1 per cent. Accurate matching of the units is important to the accuracy of the synchro system. Accuracy of the total capacitance is less important, and the total capacitance is kept to within 10 per cent of the rated value.

Synchro capacitors should never be connected in parallel with stator windings of a synchro motor and generator. When the following error of

[1] A typical size 5 Navy differential generator has an exciting current of approximately 1.0 amp, of which 0.30 amp is loss current and 0.90 amp is magnetizing current. A typical size 1 Navy control transformer requires approximately 45 ma exciting current consisting of 20 ma loss current and 40 ma magnetizing current.

the motor is nearly zero, the current between stator windings is very small; the addition of a capacitor would increase the current and reduce the accuracy of the system.

10. The Electrical-zero Autotransformer. In some synchro systems, a need arises for an electrical-zero signal, that is, for a set of voltages equal to the stator voltages of a synchro generator turned to its electrical-zero position. Instead of a properly aligned generator, an electrical-zero autotransformer may be used to supply the signal. Consider a system using standard Navy synchros. Reference to Fig. 5 and the curves of Fig. 6 shows that, when the generator is in the electrical-zero position, the voltage between stator terminals $S1$ and $S3$ is zero and the voltage of terminal $S2$ relative to $S1$ or $S3$ is 90 sin 60° or 78 volts in phase with the rotor voltage (at $R1$ relative to $R2$). These same voltages can be obtained from the $115/78$-volt autotransformer of Fig. 24.

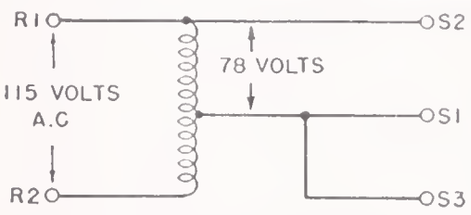


FIG. 24. Electrical-zero autotransformer.

In use, terminals $S1$, $S2$, and $S3$ are connected to appropriate stator terminals of a synchro motor, control transformer, or differential motor, and the autotransformer is a source of an electrical-zero positioning signal. Electrical-zero transformers are required for the normal operation of some synchro systems and are also employed in maintenance work for the alignment of synchro units.

11. Geared Synchro Systems. In a synchro system such as that of Fig. 1, in which the synchros rotate at the same speed as the input and output shafts, the system accuracy is limited by the no-load error of the synchros plus the (usually greater) errors caused by load torques. The amount of the error depends upon the amount of torque required and upon the steepness of the torque-angle curve near the origin (Fig. 14, p. 287).

To improve the accuracy of the systems, gear ratios are often chosen that make the synchros turn faster than the input and output shafts. For example, in Fig. 1, the synchro generator might be geared to turn at 10 times the antenna-shaft speed. The motor would then follow the generator, and the indicator coil would be geared to turn at one-tenth of the motor speed. Such gearing reduces the effect of synchro no-load errors by a factor equal to the gear ratio. Load errors are reduced even more because torque at the load shaft causes less torque on the synchro motor and a smaller error at the synchro motor shaft. To the extent that the torque-angle curve is linear, load errors are reduced by the square of the gear ratio. Of course, some additional errors are introduced by gearing because of backlash and friction.

A standard gear ratio used for Navy synchro systems is 1:36; that is, 1

revolution of the input shaft turns the generator rotor through 36 revolutions, and 36 revolutions of the rotor of the motor or control transformer correspond to 1 revolution of the output shaft. Such a synchro system is called a 36-speed system.

A disadvantage of geared systems is that the self-synchronous feature of a 1-speed system is lost. If power to a synchro system is interrupted, the generator may be turned without the motor following. With a 1-speed system, however, the motor aligns itself correctly as soon as power is restored. (If a displacement of 180 deg occurs, the motor may remain for a short time in its displaced position. However, the slightest motion of either machine will cause torques that turn the motor to its correct position.) With a 36-speed system, 36 positions of the output

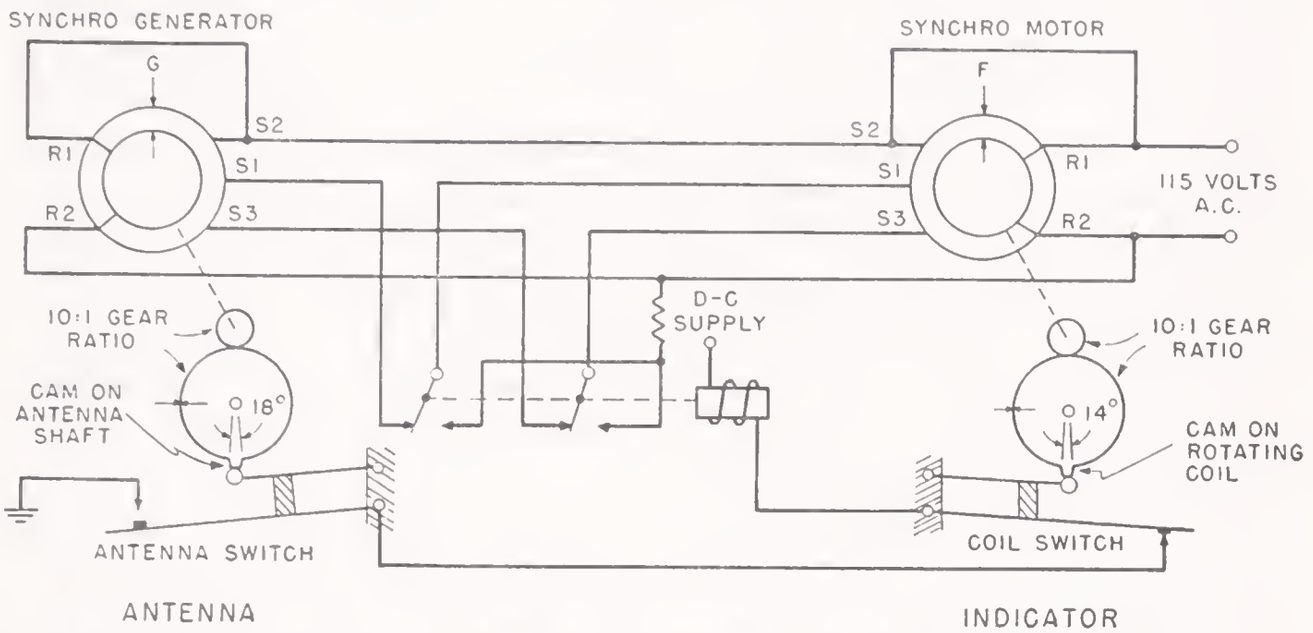


FIG. 25. Ten-speed synchro system employing cam-operated switches and used to turn deflection coil of plan-position indicator.

shaft exist for which the synchro motor is aligned with the generator. Thus any one of these positions may be assumed by the output shaft when power is restored after an interruption or when the motor is pulled out of step by a momentary high load torque.

One method of obtaining accuracy without losing self-synchronous operation employs *coarse* and *fine* synchro systems connecting the same input and output shafts. For example, 1- and 36-speed systems may be used. Incorrect alignment of the 36-speed system is prevented because, when the shaft error is large, the 1-speed system has control and reduces the error to a small value. The 36-speed system then takes control and maintains high accuracy.

A simpler system employs cam-operated switches to prevent incorrect alignment. The diagram of a 10-speed system employing such switches and used to rotate a deflection coil in correspondence with a radar antenna is shown in Fig. 25. As long as the coil is correctly aligned with

the antenna, the relay is never energized, because the switch operated by the coil cam is closed for only 14 deg of the coil rotation and the switch operated by the antenna cam is open during rotation through the same angle, plus 2 deg to either side. With the relay unenergized, the synchro generator and motor are connected in normal fashion.

As an example of the self-synchronizing action of the system, suppose that, when power is applied, the antenna is in its zero position (arrows aligned) but the coil is 180 deg away from the zero position. Both cam-operated switches are open, with the result that the synchro motor is connected to the generator and the synchros are aligned. Turning the antenna clockwise makes the indicator coil follow 180 deg out of position until the coil cam closes its switch. The relay is then energized, and instead of the generator stator voltages, an electrical-zero signal is applied to the motor. (A series resistor is used instead of an electrical-zero auto-transformer.) Therefore, the synchro motor and the indicator coil remain locked in the zero position until the antenna is turned close to its zero position. As soon as the antenna cam opens its switch, normal operation is begun.

12. Applications of Synchros. Synchros are used in radar sets (1) as the receiving and transmitting elements of remote-indication systems, (2) as variable-ratio transformers in indicator circuits, and (3) as error-detecting elements in servomechanisms. The first two applications are discussed in this article, and the third in Art. 13.

Synchro remote-indication systems are employed not only to rotate deflection coils in type P indicators, as illustrated in Figs. 1 and 25, but also in a variety of other applications to turn dials and pointers so that motions of a shaft in one location may be known at another location. Radar readings of range, azimuth, and elevation may be transmitted by synchros to the place where the data are to be used. In ships and aircraft, plan-position indicators are often made to show true bearing instead of relative bearing. That is, the indicator display shows objects in relation to the compass directions instead of relative to the heading of the ship. Such a display remains unchanged when the ship changes course. To obtain a true-bearing indicator, a differential generator is added to the generator-motor system connecting antenna and deflection coil, and the rotor of the differential is connected mechanically to a gyro-compass.

A synchro generator may be considered as a transformer in which the rotor winding is the primary and the stator windings are secondaries. The voltage ratio between primary and each secondary is changed when the rotor is turned. Synchros or synchro-like devices used to produce voltages that are functions of angular position are often called *rotary transformers*. Rotary transformers are used, for example, instead of

azimuth potentiometers in some Type B and most fixed-coil Type P indicators. Synchros have the advantage of longer life than potentiometers because the resistance wire with its sliding contact is eliminated.

In a Type B indicator, a synchro generator or control transformer may replace the azimuth potentiometer (see Art. 19, Chap. IV), provided the angle of the sector displayed is relatively small. An alternating voltage is applied to the rotor, which is made to turn with the antenna, and an output alternating voltage is obtained from one of the stator windings. The synchro is adjusted so that the output is zero when the antenna is in the zero-azimuth position. The output voltage then has one phase when the antenna is rotated clockwise from zero and a 180-deg different phase for antenna positions on the opposite side of zero. This voltage is applied to a special type of rectifier that is sensitive to phase (see Art. 16). Alternating voltage of one phase causes a positive direct voltage at the rectifier output, and voltage of the opposite phase causes a negative direct voltage. The direct voltage actuates an amplifier that supplies current to the deflection coil. A disadvantage of this system is that the synchro output voltage is a sinusoidal function of the antenna position. If the rectifier and amplifier are linear, therefore, the azimuth scale of the Type B indicator is nonlinear—unless the sector displayed is small so that only the nearly linear part of the sine wave is used.

In most fixed-coil Type P indicators, a rotary transformer and a single trapezoidal-voltage generator are employed instead of the special potentiometer and four trapezoidal-voltage generators described in Art. 22 of Chap. IV. In one form of the circuit, the trapezoidal waveform of voltage is converted by the rotary transformer into four trapezoidal driver-tube grid-voltage waves having amplitudes that vary with antenna position. In another form of circuit, the output voltage of the driver tubes is applied to the rotary transformer, which converts a single saw-tooth current wave into two waves suitable for application to the deflection coils.

A circuit of the first type is drawn in part in Fig. 26. The driver tubes and deflection coils may be connected in the same way as in the potentiometer circuit of Fig. 41, p. 273. Tube V_{14} in Fig. 26 is the switch tube of a trapezoidal-voltage generator, and V_{13} is a power amplifier for the trapezoidal wave. Power amplification is needed in order that the rotary-transformer exciting current may be supplied without distortion of the trapezoidal-voltage waveform. The rotary transformer, which is connected mechanically to the antenna, differs slightly from the usual synchro. The stator carries two windings having axes perpendicular to each other, instead of three windings with axes 120 deg apart. The amplitudes of the trapezoidal voltages obtained from the stator windings vary sinusoidally with the antenna angle, and a 90-deg phase displacement exists between the voltage amplitudes obtained from the two wind-

ings. Thus a trapezoidal waveform of voltage is present across each of the four resistors R_1, R_2, R_3, and R_4, and when the antenna turns, the four waves vary in amplitude in the correct manner for use as driver-tube grid voltages.

The stator voltages are not applied directly to the grids because neither the coupling capacitor C nor the rotary transformer can transmit the direct component of the waveform generated. As a stator voltage changes in amplitude, the base line from which the trapezoid is drawn rises or falls so that the average value of each cycle is zero. Correct

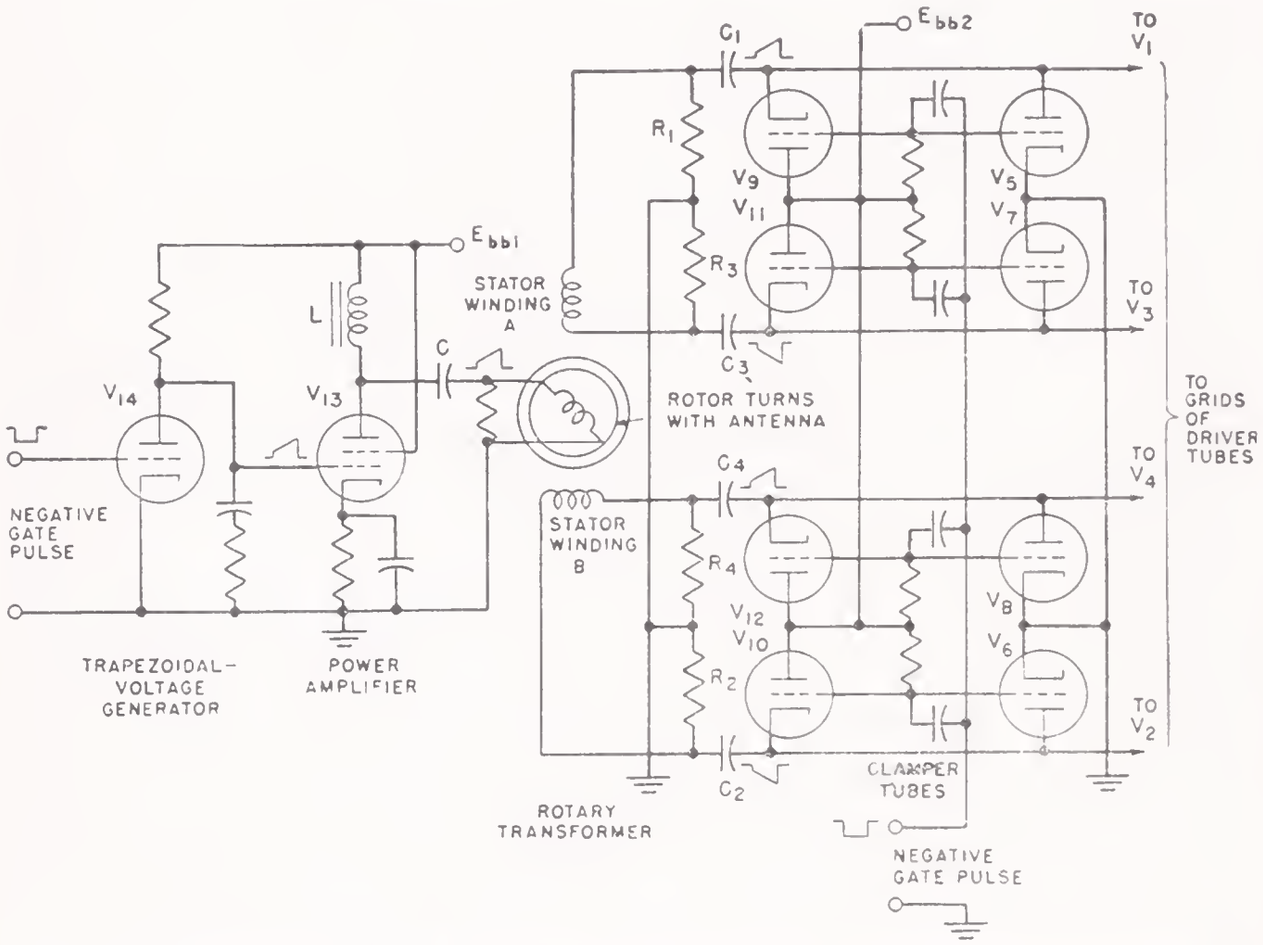


FIG. 26. Rotary transformer in sweep circuit for fixed-coil plan-position indicator.

operation of the driver tubes requires that the trapezoids be drawn up or down from a fixed base line. Therefore, the capacitors C_1, C_2, C_3, and C_4 and the clamper tubes V_5 through V_{12} are used. The circuit of the clamper tubes is like that of the switch tubes in Fig. 41, p. 273. Between sweeps, the tubes conduct and maintain the driver-tube grid voltages at the desired values. The capacitors receive whatever charges are necessary to permit these grid voltages to exist. During the sweep, the clamper tubes are cutoff by the negative grid-voltage pulses, and the trapezoidal waveforms from the rotary transformer are transferred through the capacitors to the driver-tube grids. The special clamper-tube circuit is required instead of simple diode clampers, because the trapezoidal waves may have either polarity. The base-line voltage,

which must be clamped, is sometimes at the positive extreme and sometimes at the negative extreme of the waveform.

Sometimes, instead of the special rotary transformer of Fig. 26, a standard three-stator-winding synchro is employed in fixed-coil Type P indicators. The resistance network shown in Fig. 27*A* is then employed to change the three stator voltages to two voltages having amplitude variations 90 deg apart. The variations of e_1, e_2, and the fraction of e_3 obtained from point P are plotted as functions of rotor angle in Fig. 27*B*.

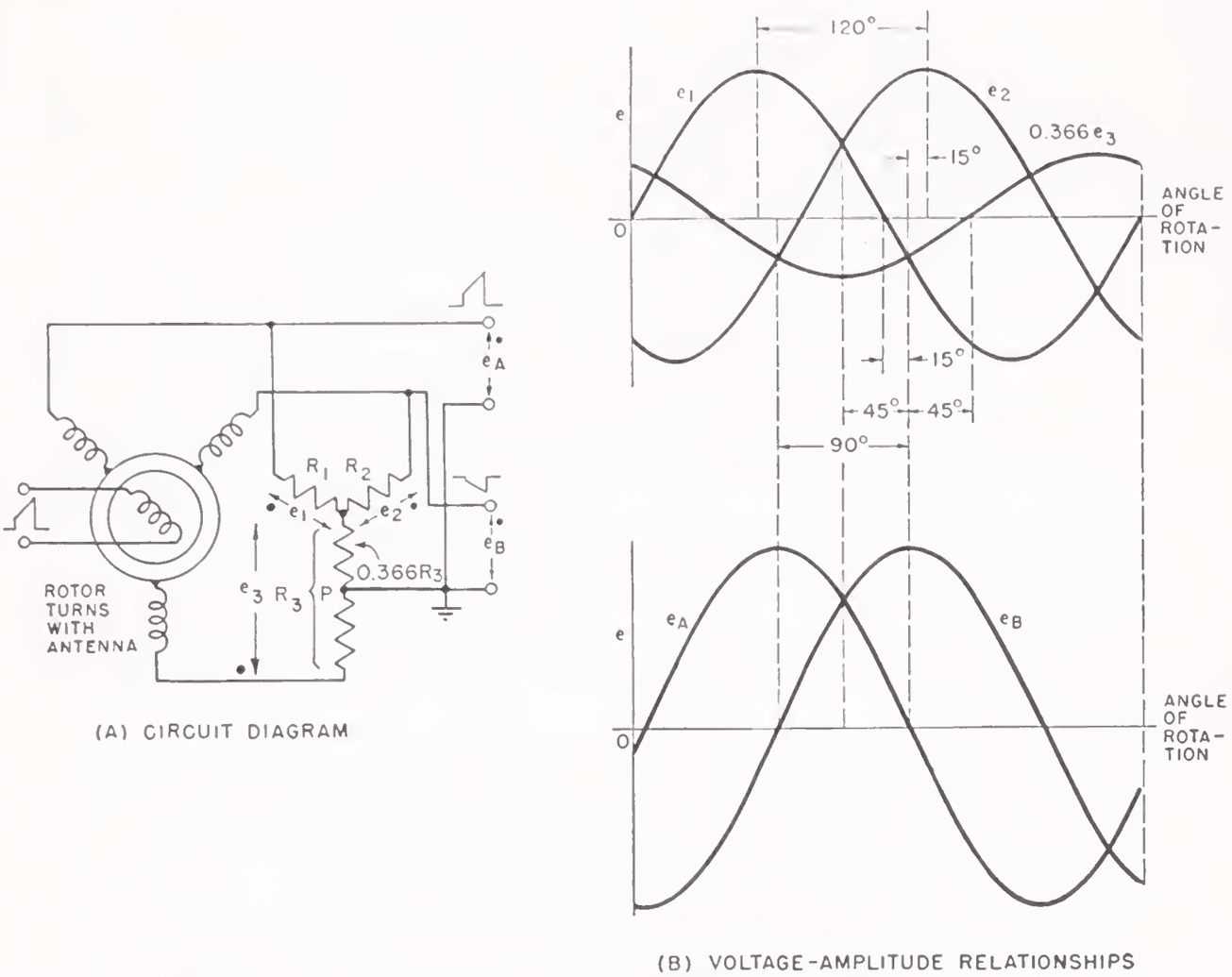


FIG. 27. Method of obtaining trapezoidal voltages with amplitude variations 90 deg apart from synchro with three stator windings.

Point-by-point subtraction of the fraction of e_3 curve from the e_1 curve yields the plot of e_A; similarly subtraction of the fraction of e_3 from e_2 yields e_B. The maximum-amplitude points of the e_1 and e_2 curves are 120 deg apart. Subtracting the fraction of e_3 causes the maximum-amplitude points of e_A and e_B to be closer together than the peaks of the e_1 and e_2 waves. The tap point P is adjusted so that the maxima of e_A and e_B are 90 deg apart.

The proper location for point P may be determined from the curves in Fig. 27*B*. As indicated in the figure, the peak of the e_B wave must occur 15 deg from the peak of the e_2 wave or 15 deg from the zero of the e_1 wave

and 45 deg from the zero of the e_3 wave. The peak of the e_B wave coincides with the zero of the e_A wave, because of the 90-deg separation of maxima, and therefore lies at a point where the ordinate of the e_1 curve equals the ordinate of the fraction of e_3 curve. Thus

$$xE_m \sin 45^\circ = E_m \sin 15^\circ \tag{1}$$

where E_m is the maximum amplitude of the e_1, e_2, and e_3 waves and xE_m is the maximum amplitude of the wave obtained from the tap point P. The value of x from Eq. 1 is sin 15°/sin 45°, or 0.366.

To produce the four driver-tube grid voltages, the e_A and e_B waves of Fig. 27 are each applied to a phase inverter (Art. 7, Chap. IV). Clamper tubes are required to restore direct components to the phase-inverter output waves; a clamping circuit like the one in Fig. 26 may be used.

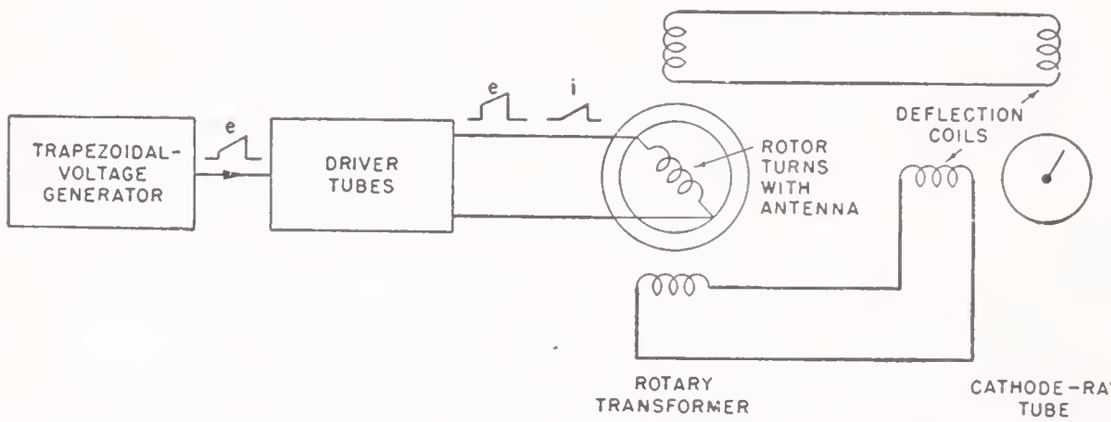


FIG. 28. Fixed-coil plan-position indicator using rotary transformer connected directly to deflection coils.

The general form of the second type of rotary-transformer circuit for plan-position indication is shown by the diagram of Fig. 28. Fixed-amplitude trapezoidal-voltage waves are applied to the grids of the driver tubes, and from these tubes trapezoidal voltage waves and saw-tooth current waves are applied through the rotary transformer to the deflection coils. The waves applied to the primary of the rotary transformer are fixed in amplitude; those delivered by the stator windings vary in amplitude with the antenna angle in the manner required to produce a rotating radial sweep.

A difficulty with the circuit of Fig. 28 is that the rotary transformer does not transmit d-c components of current, and therefore, as in the circuit of Fig. 26, a means of starting each saw-tooth sweep of the current wave from zero is needed. Clamper tubes are not readily used in the low-impedance deflection-coil circuit; thus other methods of centering the sweep are employed. In one method, the indicator spot follows a line from near one side of the screen through the center point and to the opposite side, and this line rotates about the center point. To obtain correct presentation, the transmitted pulse is delayed after the start of

the sweep until the indicator spot reaches the center of the screen, and the indicator tube is blanked during the portion of the sweep before the transmitted pulse. In another method, each saw-tooth current pulse generated in the driver tubes is followed by a rectangular pulse of opposite polarity having the same area as the saw-tooth wave. Because the saw-tooth and rectangular areas balance, the zero axes of the rotary-transformer output waves coincide with the base line from which the saw-tooth and pulse waveforms rise, and the sweep begins at the center of the indicator screen.

The rotary transformer, like a pulse transformer (Art. 20, Chap. II), distorts the waves of voltages it transmits because of leakage inductances, magnetizing inductance, and winding capacitances. Because of the air gap between rotor and stator, the ratio of magnetizing inductance to leakage inductance for a rotary transformer cannot be made as high as for a fixed-ratio transformer. Consequently, appreciable distortion of the sweep-voltage waveforms by the rotary transformer is to be expected. The nonlinearities of the range sweep that result do not materially decrease the usefulness of plan-position indicators because auxiliary circuits are always employed for accurate ranging.

A rotary transformer having two stator windings with perpendicular axes may be thought of as a special computing device for resolving a vector into components along perpendicular lines in the plane of the vector. The voltage E supplied to the rotor is made proportional to the length of the vector, and the rotor angle θ is adjusted to indicate the direction of the vector relative to perpendicular coordinate axes. The stator-coil voltages are then $E \sin \theta$ and $E \cos \theta$ and indicate the vector components parallel to the coordinate axes. In the fixed-coil Type P indicator, the rotary transformer may be considered to be such a computing device. The vector resolved is the displacement vector for the indicator spot, for which length is proportional instant by instant to range and direction is specified by the antenna azimuth angle. Rotary transformers are used in a similar way in the height-range indicator (see Fig. 6, p. 226), in which height is plotted vertically and range horizontally on the indicator screen. From the range and elevation *angle*, the height, which is range times the sine of the elevation angle, can be computed by means of a rotary transformer. In many other radar-indicator circuits and in computing circuits associated with radar equipment, rotary transformers are used to determine the components of vector quantities.

B. SERVOMECHANISMS

One means of causing an output shaft to follow the arbitrary motion of an input shaft is the synchro generator-motor system of Art. 1. Another

means is the *servomechanism*[1] or *servo*. Synchro systems can operate between shafts separated by a considerable distance but cannot supply torque amplification—the torque delivered to the load can never exceed the input torque. For this reason and because the error angle increases when large torques are transmitted, synchro systems are employed only to turn dials and pointers, move control valves, and actuate other low-torque loads. Servo systems, on the other hand, can supply the large torques required to move heavy loads, such as radar antennas, and only a very small torque need be applied to the input shaft. Remote operation is not inherent in the servo system but may be obtained if data-transmission devices (usually synchros) are made part of the servo system.

13. Components of Servo Systems. A servomechanism comprises an *error indicator* and a *controller* connected to the input and output shafts in

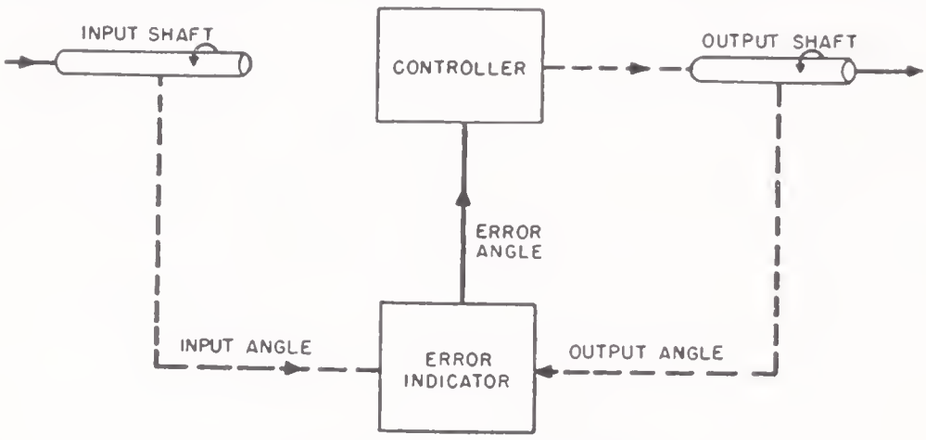


FIG. 29. Basic servo system.

the manner shown in Fig. 29. The object of the servomechanism is to cause the output shaft to repeat the motion of the input shaft by maintaining the *error angle* (deviation in angular position of output shaft from that of input shaft) as near to zero as possible. The error indicator determines the magnitude and direction of the error angle. Under control of the signal from the error indicator, the controller exerts a torque on the output shaft in a direction to reduce the error. The servo is a *closed-cycle* or *feedback* system, because a signal applied to the controller causes rotation of the output shaft and thus changes the error angle with the result that an additional signal is applied to the controller.

The error indicator and controller may take a wide variety of forms. The controller must include a *servo motor* or device for developing the output torque. Most radar servos use electric servo motors, but in associated equipment hydraulic and pneumatic devices are often employed. Servos are classified as electrical, hydraulic, pneumatic, or mechanical in

[1] H. L. HAZEN, "Theory of Servomechanisms," *J.F.I.*, **218** (September, 1934), 279–331; H. M. JAMES, N. B. NICHOLS, and R. S. PHILLIPS, *Theory of Servomechanisms* (McGraw-Hill Book Company, Inc., New York, 1947); G. S. BROWN and D. P. CAMPBELL, *Principles of Servomechanisms* (John Wiley & Sons, Inc., New York, 1948).

accordance with the type of servo motor used. The controller often contains, in addition to the motor, a power amplifier to enable the weak signal from the error indicator to control the large amounts of power supplied to the motor. The error indicator is most frequently a synchro. In electrical servos, a synchro generator and control transformer are connected mechanically to the input and output shafts. The control transformer is then the error indicator, its rotor voltage serving as input signal for the controller. In hydraulic or pneumatic servos, synchro generators may be driven by the input and output shafts and the two generators connected electrically to a differential motor. The motion of the rotor of the differential is an indication of the error and may be used to operate the control valve of the servo motor. Alternately, a control transformer

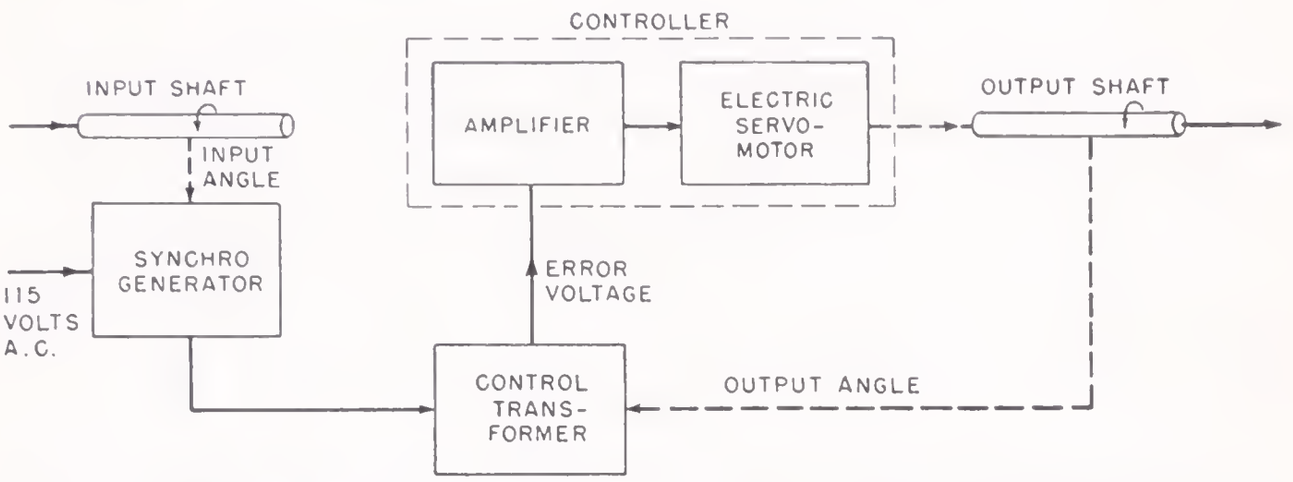


FIG. 30. Block diagram of an electrical servo using a control transformer as an error detector.

may be used as in an electrical servo, the rotor voltage being amplified and used to operate a magnetic device that controls the valve.

The most used servo in radar applications is the electrical servo with control-transformer error indicator. A block diagram of this servo is drawn in Fig. 30. If a d-c motor is used in this system, the amplifier in the figure must include a means of rectifying the alternating voltage from the synchro together with a means of increasing the power level. If an a-c motor is used, an a-c amplifier is required.

The amplifier of Fig. 30 is not necessarily an electronic device. In very low-power servos, vacuum tubes are used to supply power to the servo motor, but in larger systems, the power required by the motor is greater than can conveniently be produced by high-vacuum tubes. Sometimes thyratrons operated as grid-controlled rectifiers are employed to supply the motor power, but more often in the large d-c motor servos, special d-c generators are used. Vacuum tubes supply small field currents for the generators, and the generator output voltages are controlled by these currents.

14. Servo Using Ward-Leonard Control System. An example of a servo using a d-c motor to turn an antenna in accord with the movements of a hand crank is given in Fig. 31. In order to supply sufficient power to operate the servo motor, the amplifier in this system comprises two d-c generators in addition to the vacuum-tube circuits of the servo-amplifier unit. If any error exists between the angular positions of the hand crank and antenna, an error voltage is applied by the synchro control transformer to the servo-amplifier. The output of the servo-amplifier is a direct field current for the small d-c generator called an exciter. The output of the exciter is a larger direct current applied to the field winding

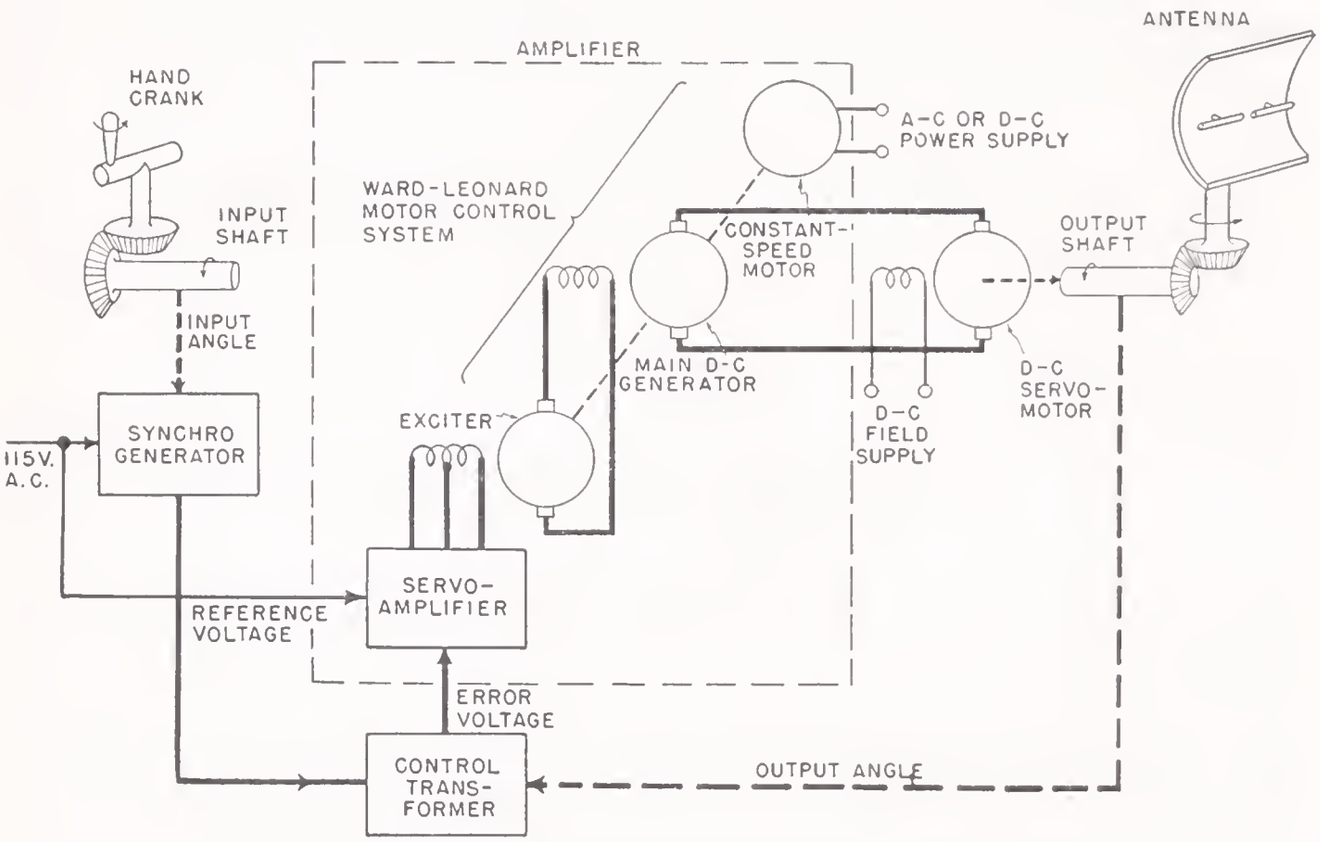


FIG. 31. Antenna-control servo using Ward-Leonard system.

of the main d-c generator. The output of the main generator is a direct current large enough to operate the servo motor. Thus existence of an error causes the motor to turn the antenna in the direction to correct the error, and the antenna is made to follow movements of the hand crank accurately.

Because the motor must be capable of turning the antenna in either direction, the output of each generator must be reversible in polarity. The servo-amplifier output current must therefore reverse in direction when the error reverses.

The Ward-Leonard connection of two d-c generators in cascade is commonly used in industry for speed control. The two generators are coupled to the same shaft and are driven by a constant-speed, a-c or d-c, motor. Because the power obtained from each generator armature is

much larger than the field power required for control, the generators may be considered as *dynamo-electric* power amplifiers. The source of the generator output power is the mechanical power supplied by the constant-speed motor, this prime mover being analogous to the plate power supply in a vacuum-tube amplifier.

15. Direct-current Servo Motor. The essential features of the d-c servo motor in Fig. 31 are shown in the cross-section drawing and schematic diagram of Fig. 32. The machine has a stationary *field* structure made of iron and a rotating armature built up from iron laminations. The field and armature each carry a winding; connections to the armature

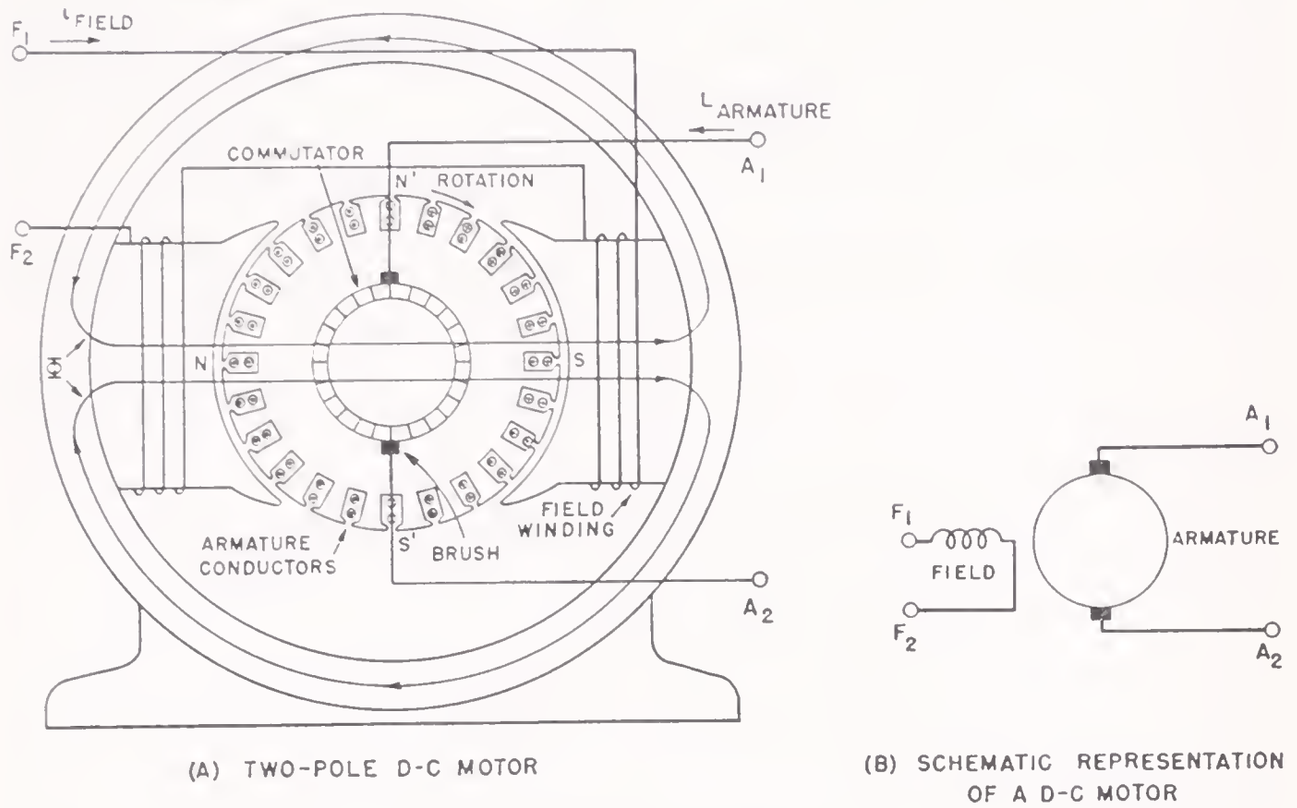


FIG. 32. D-c motor.

winding are made through brushes riding on the *commutator*, which is a copper ring separated into segments insulated from one another.

If a current is sent through the field winding in the direction of the i_{field} arrow, a magnetic flux is established along the paths indicated by Φ, and north and south field poles are produced at the positions marked N and S.

The armature winding is arranged as a coil with its axis perpendicular to the field-flux lines. When the armature-terminal current is in the direction of the i_{armature} arrow, current in the dot-marked armature conductors is directed out of the page and current in the cross-marked conductors is directed into the page. These currents magnetize the armature, producing the poles N′ and S′.

The magnetized armature turns in a clockwise direction, because it tends to align itself with the field flux. As the armature turns, the com-

mutator acts as an automatic switch to reverse currents in the armature conductors, one at a time, and to maintain constant the location of armature currents with respect to the field poles. Thus continuous torque is obtained during rotation. For constant field strength, the torque produced is proportional to the armature current and reverses in direction when the armature current is reversed. The direction of the torque is also reversed if the field current is reversed without changing the direction of the armature current.

Motion of the armature conductors through the field flux induces in them a voltage called the *armature induced voltage* or *counter electromotive force*. The induced voltage is in a direction to oppose the flow of armature current and has a magnitude proportional to the product of field flux and speed of rotation. The impressed voltage at the armature terminals exceeds the induced voltage by the voltage drop produced by the armature current flowing through the armature resistance. That is,

Terminal voltage = induced voltage + armature-resistance voltage drop

or

$$e_t = k\Phi n + i_a R_a \tag{2}$$

where

e_t is the armature terminal voltage,
k is a constant of proportionality,
Φ is the field flux,
n is the motor speed,
i_a is the armature current,
R_a is the armature resistance.

The armature resistance is low—only a small fraction of an ohm in large machines—and in operation at normal speeds the resistive voltage drop is so small that the armature terminal voltage very nearly equals the induced voltage.

Because the armature resistance is low, the field flux and armature terminal voltage determine the speed of rotation. Application of a mechanical load tends to slow the motor and to reduce the induced voltage. As a result, a larger armature current flows and an increased torque is available to accelerate the load and maintain the induced voltage nearly equal to the terminal voltage. To show the effect of armature voltage and field flux on speed, Eq. 2 may be rewritten as

$$n = \frac{e_t - i_a R_a}{k\Phi} \tag{3}$$

According to Eq. 3, the motor speed is very nearly proportional to the armature voltage if the field flux is maintained constant. If the armature

voltage is held constant, the speed varies inversely with the flux; that is, an *increase* of field current causes a *decrease* of speed.

In the servo system of Fig. 31, the *armature* voltage of the servo motor is controlled by the Ward-Leonard system, and the field current is maintained constant. This system is rugged, can supply large amounts of power, and can produce large torques to change the load speed quickly. It has the disadvantage that the large armature power of the motor is controlled instead of the much smaller field power.

Another system employed with relatively low-power servos uses a vacuum-tube servo-amplifier to control the field current of a d-c motor. The armature cannot then be supplied with a constant voltage, because it would become essentially a short circuit across the line each time the field current is reduced to zero. Instead, special circuits are used which supply a nearly constant current to the armature with the result that a torque approximately proportional to field current is obtained.

Sometimes low-power servos employ d-c motors with permanent-magnet fields. The armature is then supplied from a dynamo-electric amplifier, such as the Ward-Leonard generators. The permanent-magnet motor avoids the necessity of a separate field-current supply, but it has the disadvantage of being somewhat larger than an ordinary motor of equal power rating. Armature reaction—that is, magnetization of the armature and parts of the field structure caused by armature current—has a tendency to demagnetize the permanent magnets. To cancel the field of the armature in the region of the permanent magnets, a *compensating winding* is placed in slots in the pole faces and is connected in series with the armature winding. Each armature conductor under a pole face has a compensating-winding conductor close to it that carries an equal current in the opposite direction. Current should never be sent through the armature of a permanent-magnet motor unless the compensating winding is connected in series with it.

Direct-current generators, for example, those in the Ward-Leonard system, are exactly like d-c motors in construction; in fact, a single d-c machine may be used as either motor or generator. If field current is supplied to a d-c machine and the armature is rotated, the armature induced voltage may be used as a source of electrical power. Except for the small resistive drop in the armature, the terminal voltage is proportional to the product of field flux and speed of rotation; that is,

$$e_t = k\Phi n - i_a R_a \tag{4}$$

Current drawn from the terminals of a generator causes a torque exactly as in a motor, and this torque places a mechanical load on the machine that drives the generator. Thus the source of the electrical power obtained from a generator is the mechanical power applied to its shaft.

16. Phase-sensitive Rectifiers. The servo-amplifier of Fig. 31 contains vacuum-tube circuits that amplify and rectify the a-c error signal from the control transformer, amplify the resulting direct current, and apply the output to the field winding of the exciter generator. The rectifier component of the servo-amplifier is called a *phase-sensitive rectifier* because it produces a direct voltage or current the polarity of which reverses when the phase of the control-transformer voltage reverses.

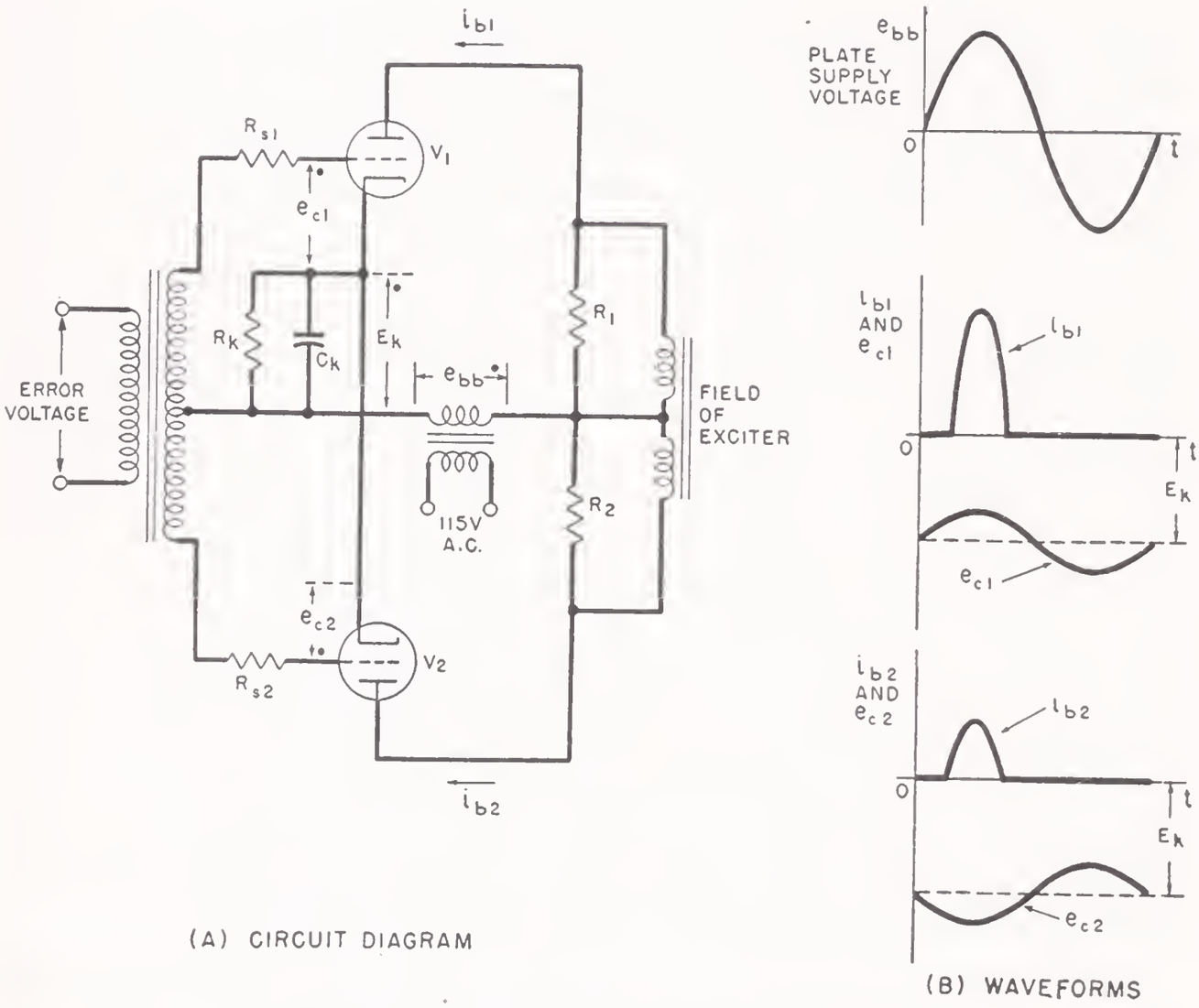


FIG. 33. Servo-amplifier for servo of Fig. 31.

The circuit of Fig. 33*A* can be the entire servo-amplifier if the power requirements of the exciter field are small. This circuit amplifies the error signal and provides phase-sensitive rectification. Two triodes are connected in push-pull and supplied with plate voltage directly from the a-c line, the two plate voltages being in phase with each other. The two grid voltages are obtained from the synchro control transformer and are 180 deg out of phase with each other. Because the synchro system is operated from the same a-c line as the servo-amplifier (see Fig. 31), each grid voltage is either in phase or 180 deg out of phase with the plate voltages.

Waveforms of grid voltage for one direction of error and of plate supply voltage are shown in Fig. 33*B*. Plate currents can flow only during the half cycle in which the plate voltages are positive; therefore pulses of current are obtained as shown in the figure. The dotted lines used as axes for the sine waves of grid voltage represent the cathode-bias voltage produced by the average plate current flowing through resistor R_k.

If the error voltage is zero, the pulses of current in V_1 and V_2 are the same. The center-tapped field winding of the exciter is connected so that the two plate currents tend to magnetize it in opposite directions. Thus no field is produced in the exciter when balanced currents flow. For an error voltage in one direction, the grid-voltage waves indicated by e_{c1} and e_{c2} in Fig. 33*B* are obtained. As a result, the current in tube V_1 is made greater than that in tube V_2 and a magnetization of the exciter field is produced by the difference of the currents. If the error signal is reversed in polarity, tube V_2 draws the greater current and the direction of the exciter field flux is reversed.

Resistors R_{s1} and R_{s2} are large resistors used to limit grid current. When the error angle is large, the peak voltage at the grid-transformer terminals is much greater than the bias voltage, and excessive grid current would result if the resistors were not present.

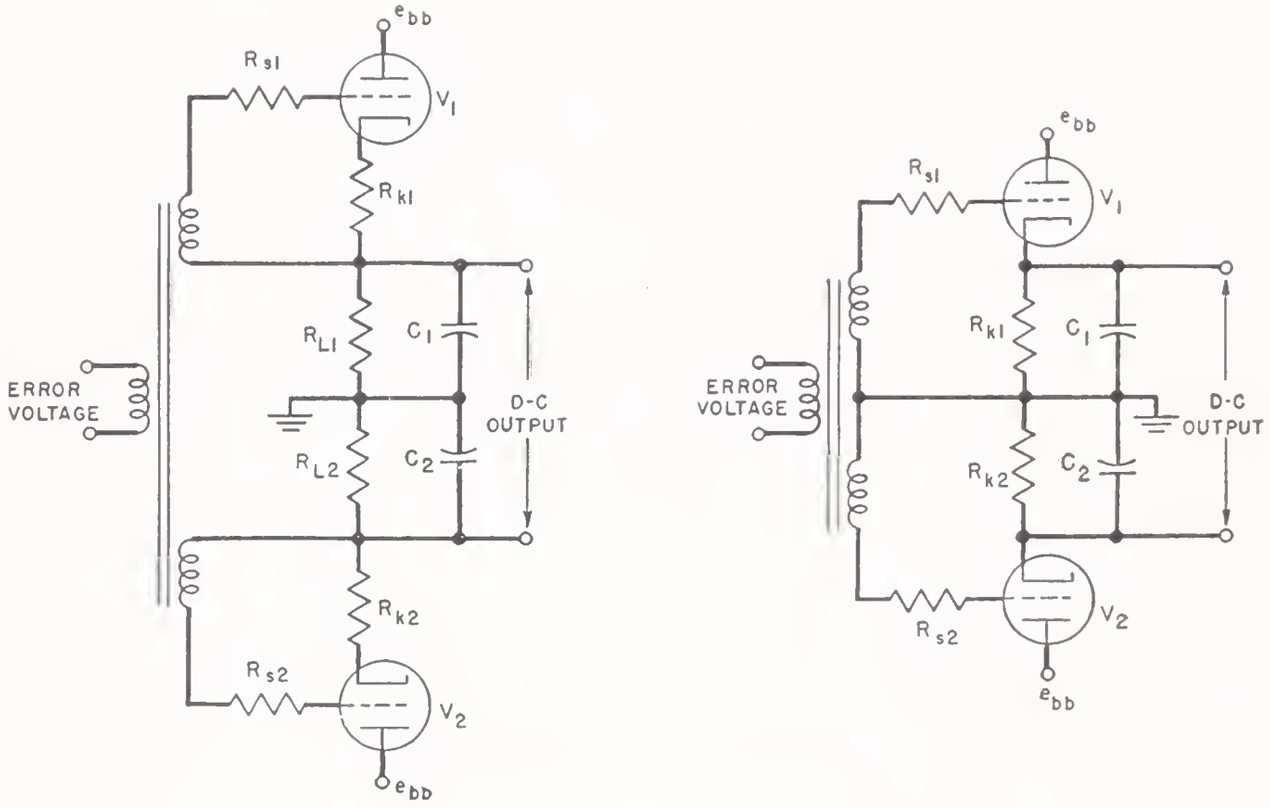
Resistors R_1 and R_2 are necessary because of the high inductance of the field winding. If these resistors were omitted, the field currents would be very small, because little current could be built up in the large inductances during the short conduction period of the tubes. With the resistors present, a little current builds up during the first conduction period, and the current increases in small steps during succeeding conduction periods —during the cutoff periods of the tubes the field currents flow through field winding and resistors. A steady-state condition is reached in which the field current is nearly steady but which increases and decreases a little as the tubes conduct and are cut off. Another reason that resistors R_1 and R_2 are required is to protect the insulation of the field windings. Without the resistors, extremely high voltages would appear across the field coil each time the tube current is suddenly cut off.

The lower the values of R_1 and R_2 are made, the more nearly steady the field currents become and the lower is the inductive-impulse voltage that can appear across the field windings. Excessively small values of R_1 and R_2, however, reduce the amount of the field current by providing a shunt path for plate current of the tubes. Furthermore, small values of R_1 and R_2 increase the delay in the response of the circuit to a sudden change in the error; the currents in the tubes change immediately, but the currents in the field-circuit inductance require several cycles of operation to reach their new values.

Frequently, filter capacitors are used instead of resistors R_1 and R_2.

The capacitors provide a similar, but somewhat greater, smoothing action and permit larger average field currents to be obtained. On the other hand, they produce a greater lag in the response of the circuit.

If somewhat more power is required by the exciter field, a push-pull d-c amplifier may be placed between the phase-sensitive rectifier and the field winding. The phase-sensitive rectifier may have the circuit of Fig. 33*A*, with the field windings replaced by parallel *R-C* loads. Relatively smooth direct voltages are then obtained across the load impedances. The direct coupling required in such a circuit makes it impossible to



(A) CATHODE-LOADED-AMPLIFIER CONNECTION (B) CATHODE-FOLLOWER CONNECTION

FIG. 34. Phase-sensitive rectifier circuits with load in the cathode circuit.

operate the cathodes of both stages near ground potential. Consequently, the circuit of Fig. 33*A* is frequently replaced by one of the cathode-loaded phase-sensitive rectifiers of Fig. 34. Operation of the circuit of Fig. 34*A* is the same as that of a plate-loaded circuit. The circuit of Fig. 34*B* is like the cathode-follower-connected detector (see Art. 27, Chap. VI). In this circuit the direct cathode-to-cathode voltage cannot exceed the peak value of the applied alternating grid-to-grid voltage.

Diodes instead of triodes may be used in phase-sensitive rectifiers as shown in the circuit of Fig. 35. The reference voltage must be obtained from the same a-c line that supplies power to the synchros. When the error is zero, both diodes rectify the reference voltage and no direct voltage appears between output terminals. Existence of an error

causes a greater voltage to be applied to one diode and a smaller voltage to the other; an unbalance of the rectifier output voltages results.

In all phase-sensitive rectifiers for servos, a compromise must be made in the choice of filter elements. Large capacitances or inductances are

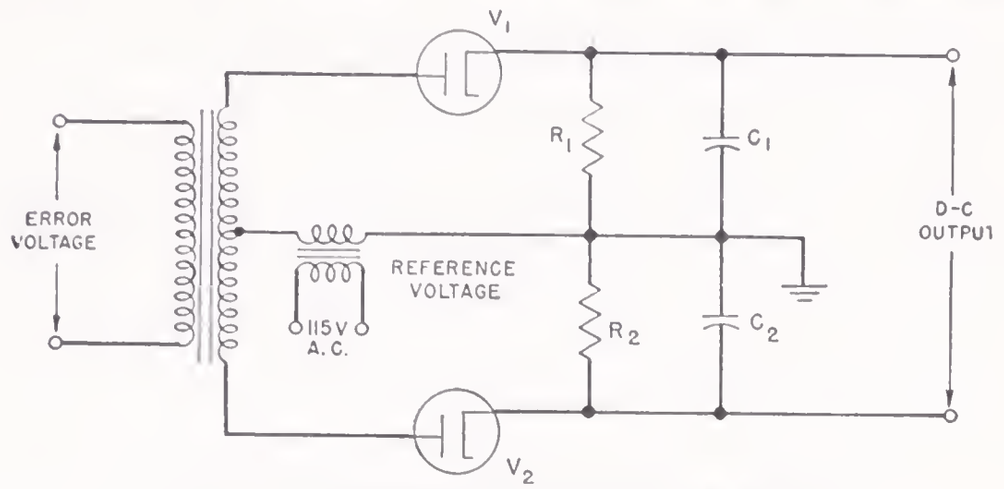


FIG. 35. Phase-sensitive rectifier using diodes.

needed to smooth the output voltage adequately, yet small elements should be used to allow the output voltage to change suddenly to follow sudden changes of the input. As explained in Art. 21, sluggish response in the controller of a servo increases the tendency of the system to oscillate.

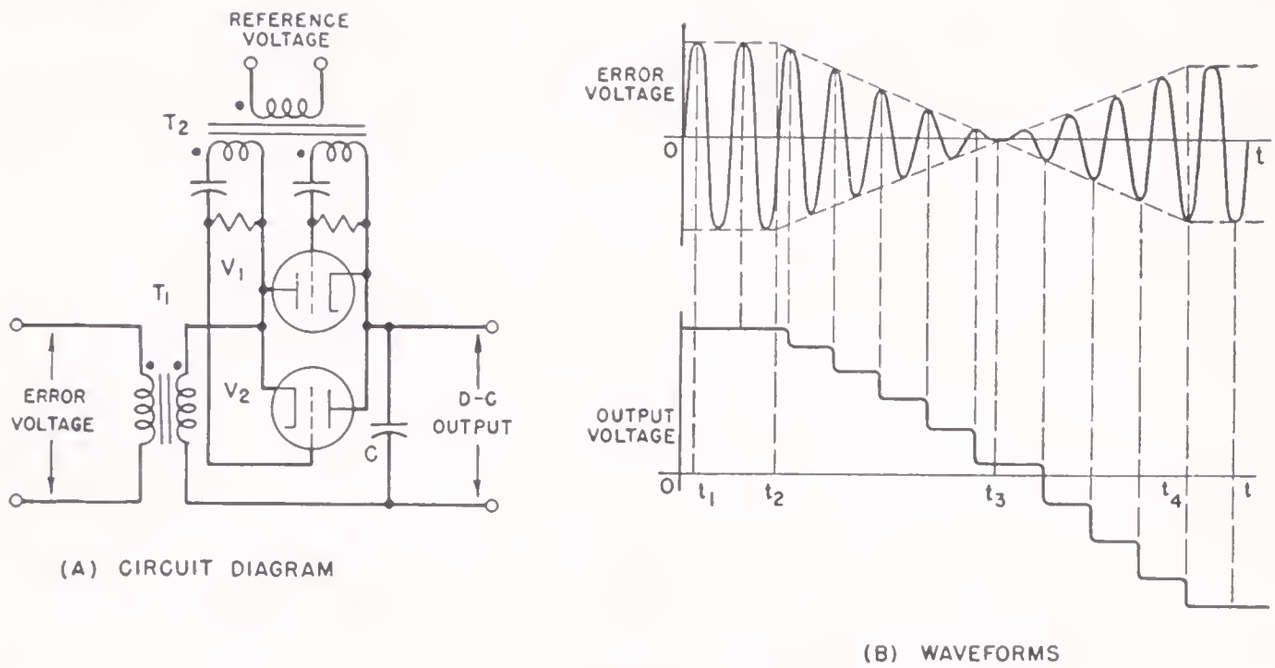


FIG. 36. Keyed rectifier.

A form of phase-sensitive rectifier that provides especially fast response with little ripple is the *keyed rectifier* of Fig. 36. In Fig. 36*A*, the reference-voltage sine wave is applied to the grids of triodes V_1 and V_2 through *R-C* grid-leak-bias circuits. A large bias voltage is produced so that the tubes are cutoff during most of the cycle. On one peak of the reference-voltage cycle, the two tubes are simultaneously capable of con-

duction. Therefore, at the conduction instants, capacitor C is charged or discharged until its voltage is nearly equal to the instantaneous error voltage from transformer T_1. Two tubes are used so that the voltage across C can be either raised or lowered by conduction in one tube or the other. After the brief conducting period for V_1 and V_2, capacitor C is isolated from the remainder of the circuit and maintains a constant voltage until the conducting period of the next cycle.

The error voltage is either in phase or 180 deg out of phase with the reference voltage. Thus the keyed-rectifier output voltage follows the peak value of the error-voltage wave in the manner illustrated by the waveforms of Fig. 36*B*. Between times t_1 and t_2, tube V_1 conducts during positive peaks of the error-voltage wave and charges C to the voltage of the peaks. In the time interval t_2 to t_3, the voltage across C decreases in steps, a change of voltage occurring once each cycle when C is discharged through V_2 to the new and lower peak value of the error voltage. At time t_3, the phase of the error voltage reverses so that negative instead of positive peaks occur at the conduction times. Therefore, the output voltage continues to decrease through negative values until a steady voltage is reached at time t_4. The keyed rectifier provides a direct voltage equal in magnitude to the peak value of the error voltage and of a polarity that reverses when the error-voltage wave reverses phase. The ripple voltage is negligible except when the error voltage is changing.

17. Amplidyne Servomechanism. Figure 37 shows the block diagram of a servomechanism that is identical with the one in Fig. 31, except that the two generators of the Ward-Leonard system are replaced by a single

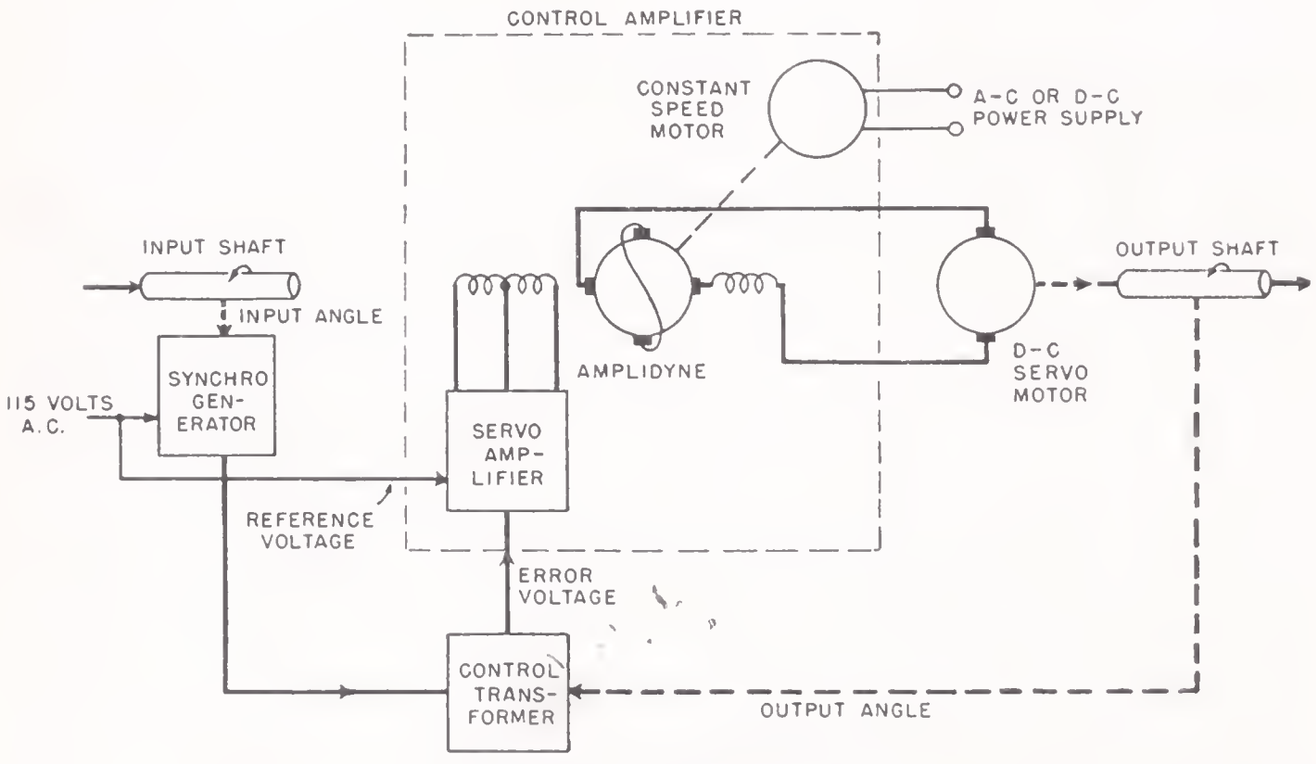
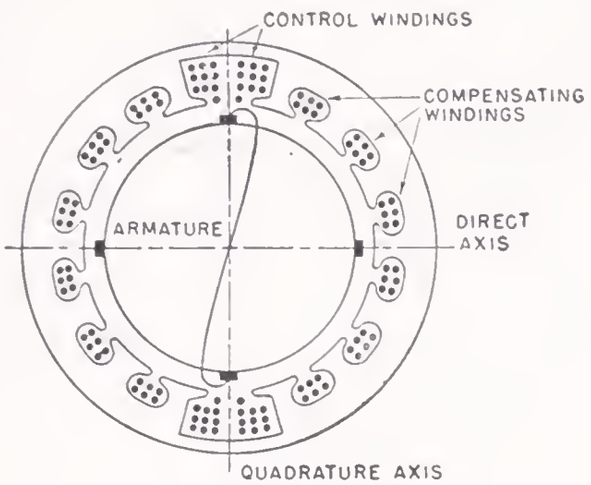


FIG. 37. Amplidyne servo.

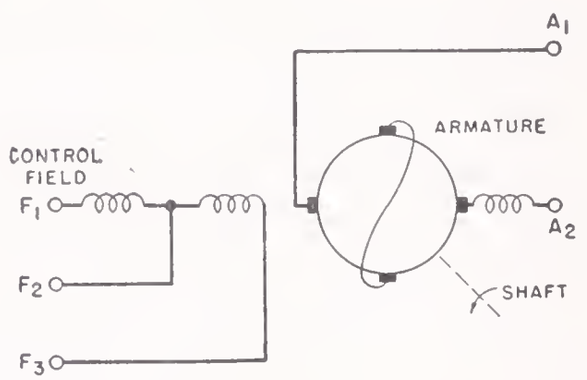
rotating machine called an *Amplidyne*. The Amplidyne is a special form of d-c generator particularly useful in servomechanisms.

The construction of an Amplidyne is shown by the cross-section drawing of Fig. 38*A*. The machine is similar to an ordinary d-c motor or generator, but the field poles are very short and wide and carry several different windings. Four brushes make contact with the commutator. Two of the brushes are in the positions of the brushes of an ordinary machine and are short-circuited. The other two brushes lie on a perpendicular axis and are the output connections for the machine. The schematic representation of an Amplidyne is shown in Fig. 38*B*. The Amplidyne generator is driven by a constant-speed motor. To reduce the size and weight of the Amplidyne and the drive motor, a high speed of rotation (1,800 to 4,000 rpm) is used.

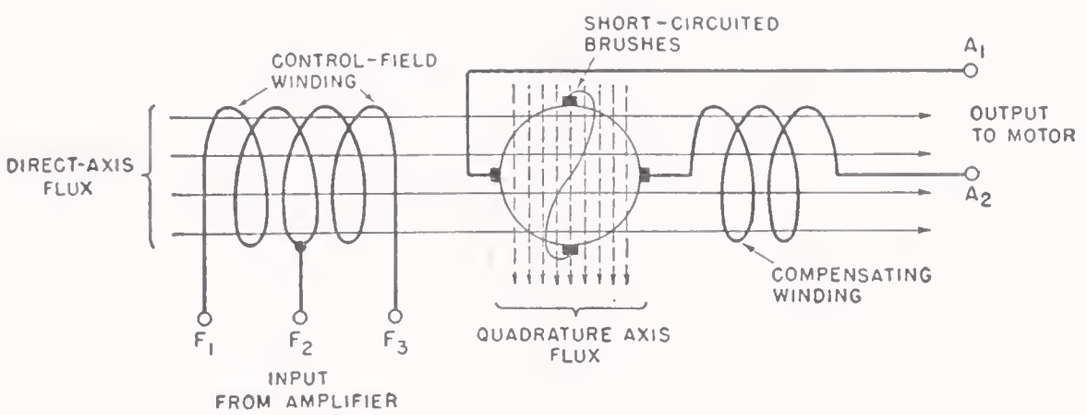
The Amplidyne may be considered as two d-c generators or dynamoelectric amplifiers combined into one. The first "generator" or "stage of amplification" includes the control-field winding and the short-circuited brushes. Current from the servo-amplifier sets up flux along the direct axis of the machine (see Figs. 38*A* and *C*). Rotation of the



(A) CROSS SECTION



(B) SCHEMATIC REPRESENTATION



(C) FLUX RELATIONS

FIG. 38. The Amplidyne generator.

armature conductors through this flux induces a voltage in the conductors that causes a current to flow through the short-circuited brushes. Because the brushes are short-circuited, only a small control-field current is needed to produce a large current in the armature conductors and short circuit. This current magnetizes the machine in a direction perpendicular to the control-field flux (as explained in the discussion of d-c motors, Art. 15) and produces the quadrature-axis flux indicated by dotted lines in Fig. 38*C*. Sometimes the brushes are connected to a low-resistance field winding on the quadrature axis, and this winding aids the armature reaction in setting up quadrature-axis flux. The first stage of amplification produces a quadrature-axis flux much greater than the original control-field flux.

The second stage of amplification makes use of the quadrature-axis flux and the output brushes. Because this flux is shifted 90 deg from the direct-axis flux, the voltage produced by rotation of the armature through the quadrature flux appears at brushes located 90 deg from the short-circuited brushes. These brushes may be used as output terminals to supply voltage and current to a load. The total amplification of power between the control field and output terminals varies from several hundred to several tens of thousands, depending upon the design of the machine.

The compensating winding shown in Figs. 38*A* and *C* is necessary in order that appreciable current may be drawn from the output terminals. Just as current in the short circuit magnetizes the machine along an axis 90 deg from the control flux, so current in the output circuit tends to magnetize the machine 90 deg from the quadrature-axis flux along the direct axis in opposition to the control-field flux. Without the compensating winding, even a small output current would reduce the direct-axis flux almost to zero. To cancel the armature reaction of the output current, a compensating winding like that used with a permanent-magnet motor is placed in slots in the pole faces and connected in series with the output brushes. The compensating winding is arranged to cancel the magnetomotive force set up by output currents, but not to affect that caused by current in the short-circuited brushes.

Armature reaction caused by output current is analogous to negative feedback in a vacuum-tube circuit. The compensating winding may be considered as positive feedback added to neutralize the degeneration. The amount of compensation must be exactly right. Undercompensation produces poor voltage regulation; overcompensation yields a net positive feedback and may cause the flux and voltage to build up until limited by magnetic saturation.

Because of the high amplification of the Amplidyne, even a small amount of residual magnetism in the direct axis of the machine can cause

an appreciable output voltage when the control-field current is zero. To eliminate the residual magnetism, a demagnetizing ("killer") winding is sometimes provided on the direct axis of the field structure. It is supplied with alternating current from a generator mounted on the same shaft as the Amplidyne.

18. Two-phase Induction Motor. Several types of a-c motors are useful for driving the output shafts of servomechanisms. The most important of these is the two-phase induction motor, which has wide application in low-power servos such as those used in computers and radar automatic-tracking and aided-tracking circuits, and those used to turn

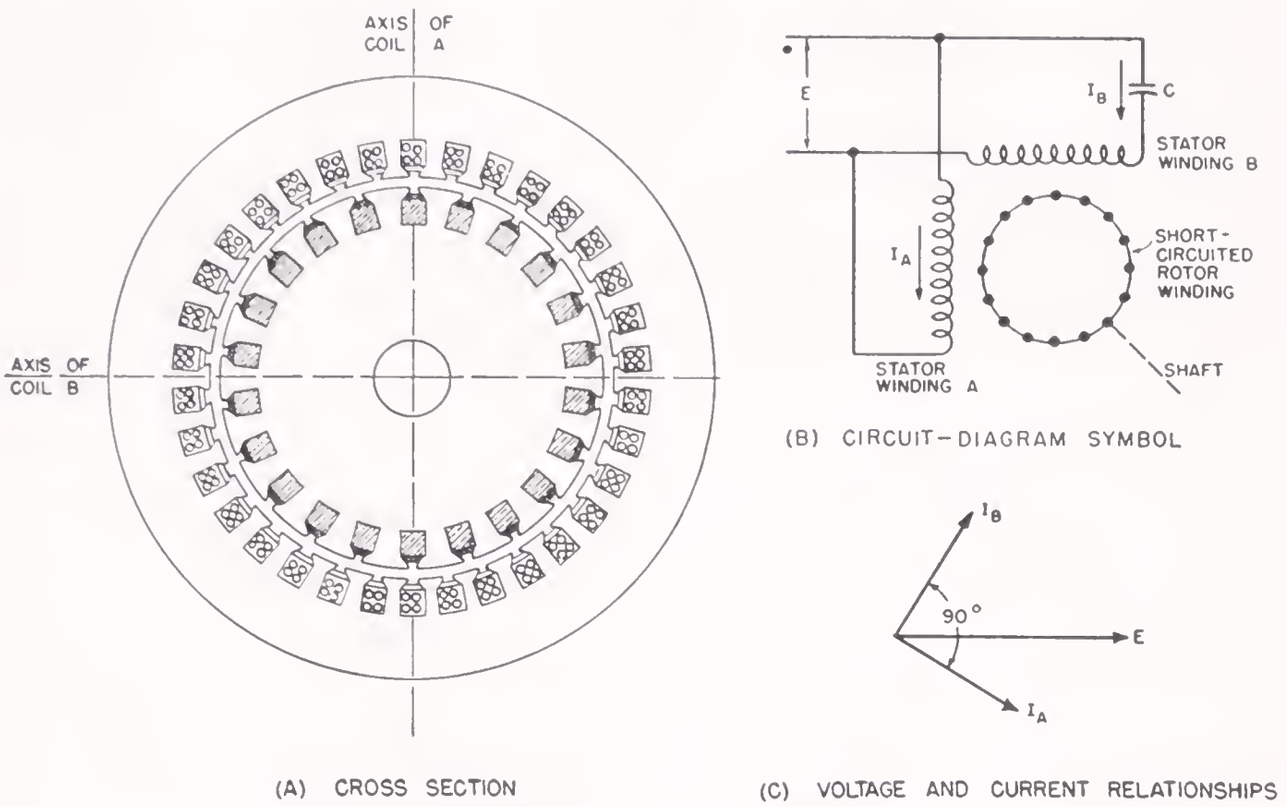


FIG. 39. Two-pole two-phase induction motor.

deflection coils of plan-position indicators. To understand servomechanisms employing induction motors, it is first necessary to know the characteristics of these motors.

Figure 39 shows a cross section and a circuit-diagram representation of a two-pole two-phase induction motor. Stator and rotor are built of sheet-steel laminations. The stator has two similar windings, coil A and coil B, arranged to magnetize the machine in perpendicular directions. The rotor may carry either a short-circuited winding of wire or a squirrel-cage winding. The squirrel-cage winding consists of conducting bars in the rotor slots, the bars being short-circuited at each end of the rotor by conducting rings.

The stator coils are usually supplied with alternating currents equal in magnitude but 90 deg apart in phase. Such currents may be obtained directly from a two-phase power system, or they may be obtained from a

single-phase source by means of a phase-shifting circuit, like that in Fig. 39*B*. As indicated in the complex-number diagram of Fig. 39*C*, the current I_A in coil *A* lags the applied voltage *E* because of the inductance of the winding. Current I_B leads *E* because capacitor *C* has a reactance greater than the inductive reactance of coil *B*. Adjustment of the size of the capacitance changes both the phase and magnitude of I_B and allows the relationship of currents in Fig. 39*C* to be approximated.

As a result of current in both coils, the total flux through the rotor core is constant in magnitude and rotates about the axis of the motor at a

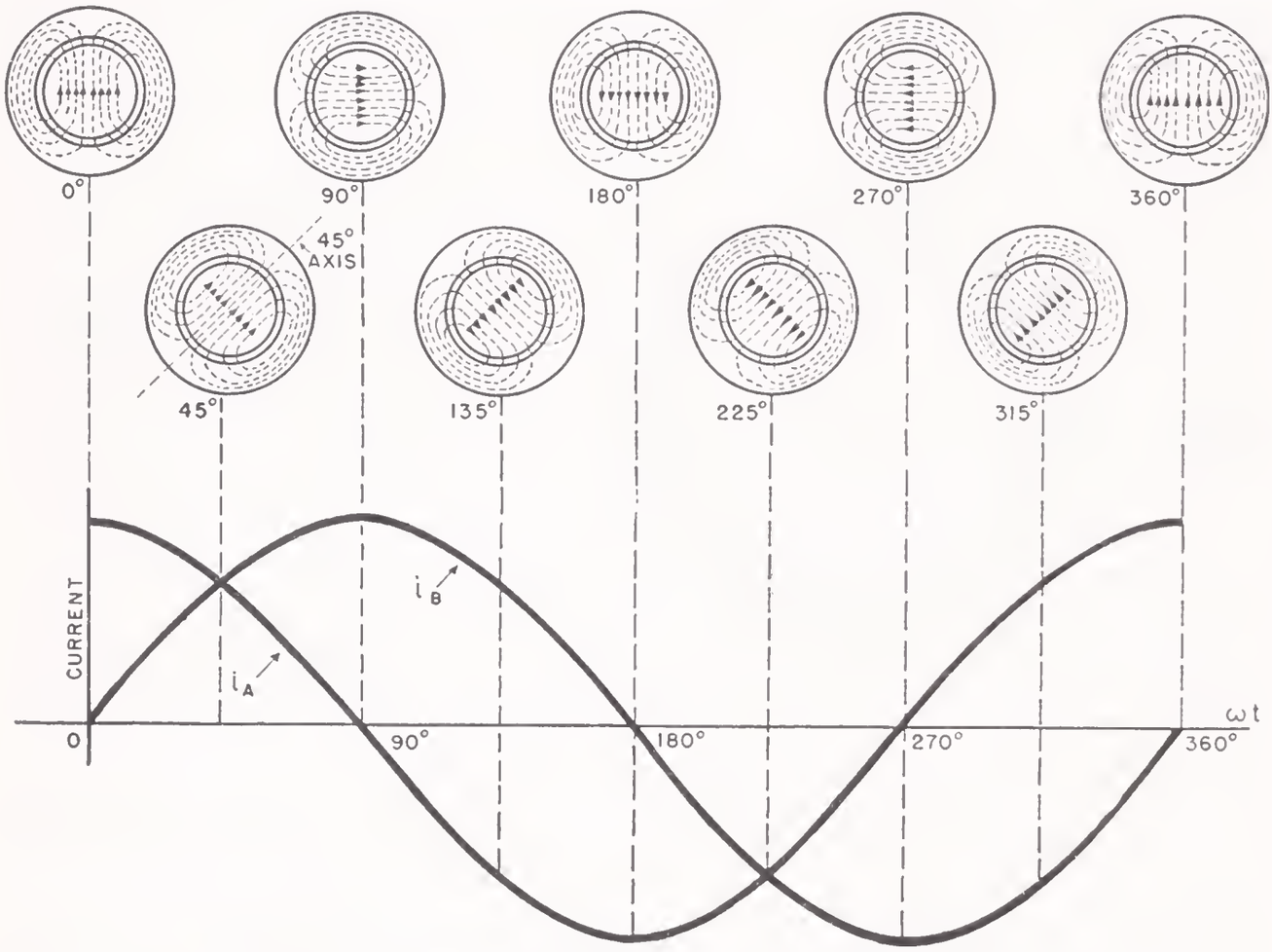


FIG. 40. Rotating field in two-phase induction motor.

speed determined by the frequency of the stator currents. The diagram of Fig. 40 shows how a rotating flux is produced. The top row of drawings in this figure shows the flux patterns in the machine at 90-deg intervals in the current cycle. When the phase angle is 0 deg, only coil *A* carries current, and the flux is directed upward along the axis of coil *A*. At 90 deg, only coil *B* carries current, and the flux is directed to the right. At 180 deg, the flux is directed downward because coil *A* carries current in the negative direction. The patterns in the top row show that the flux rotates one revolution for each cycle of the supply current.

Flux patterns obtained at intermediate values of phase angle, when current flows in both coils, are given in the second row of Fig. 40. Because the turns of the stator coils are distributed in the slots in such a way that

the air-gap flux density varies sinusoidally with angle around the rotor, the flux remains constant in magnitude throughout its rotation, and the speed of rotation is constant. The relationships are similar to those used to produce the circular sweep of a Type J indicator (see Art. 10, Chap. IV) or a rotating range trace in a fixed-coil plan-position indicator (see Art. 22, Chap. IV).

The speed at which the flux rotates is called the *synchronous speed* of the machine. For a two-pole motor operated from 60-cps power, the synchronous speed is 60 rps or 3,600 rpm. The flux always turns past one pair of poles during each cycle of the stator currents. In a four-pole machine, for example, two cycles of supply current are required for each revolution, and the synchronous speed is 1,800 rpm when 60-cps power is used.

The *direction* of rotation of the flux is reversed if the terminal connections of *either* stator coil are interchanged. To see that this is true, redraw Fig. 40 with i_A or i_B (but not both) reversed in polarity.

Rotation of the magnetic field past the conductors of the rotor induces voltages in these conductors, and currents therefore flow in the short-circuited rotor winding. The directions of rotor currents are indicated in Fig. 41 by dots and crosses for a particular instant during the rotation of the flux. (In the figure, rotor currents are assumed to be in phase with the rotor induced voltage. This assumption is correct only when the rotor speed nearly equals the synchronous speed.) The letters N and S indicate the rotating north and south poles of the stator, and N′ and S′ represent the north and south poles of the rotor produced by rotor currents. The rotor is thus a magnet that tends to align itself with the stator field; a torque is produced in a direction to make the rotor follow the rotating magnetic field of the stator.

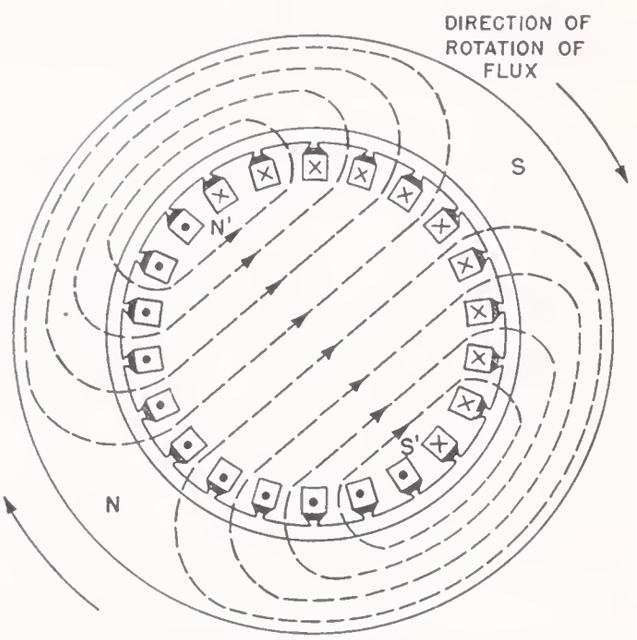


FIG. 41. Relation of rotating flux to rotor currents and torque in two-phase induction motor.

The speed of rotation of an induction motor is always less than the synchronous speed, because relative motion between rotor and flux is needed in order for voltages to be induced in the rotor winding. Typical curves of torque produced as a function of speed are shown in Fig. 42. The torque varies with rotor speed because both the magnitude and the frequency of the induced voltage depend upon the speed of the rotor

relative to the speed of the flux. Changes of rotor frequency change the resistance and reactance of the winding and the phase angle of the rotor current; changes of the phase angle of the rotor current shift the rotor poles N′ and S′ relative to the stator poles N and S.

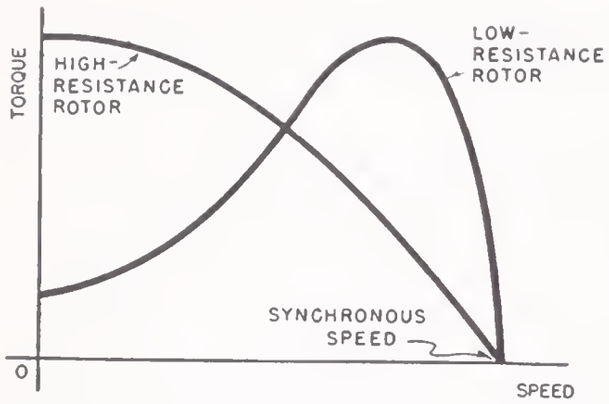


FIG. 42. Torque-speed curves of induction motor.

Induction motors commonly used in applications other than servomechanisms have low-resistance rotors. These motors operate at a speed little affected by the load torque (see Fig. 42) and at high efficiency. Motors with high-resistance rotors, however, are necessary in servomechanisms in order that the motor speed can be controlled smoothly. The maximum torque should occur at zero speed in a servo motor, in order to prevent a momentary increase in speed from producing excess torque and a tendency to overshoot and oscillate.

19. Servomechanisms Using Two-phase Induction Motor. The usual method of controlling an induction motor in a low-power servomechanism

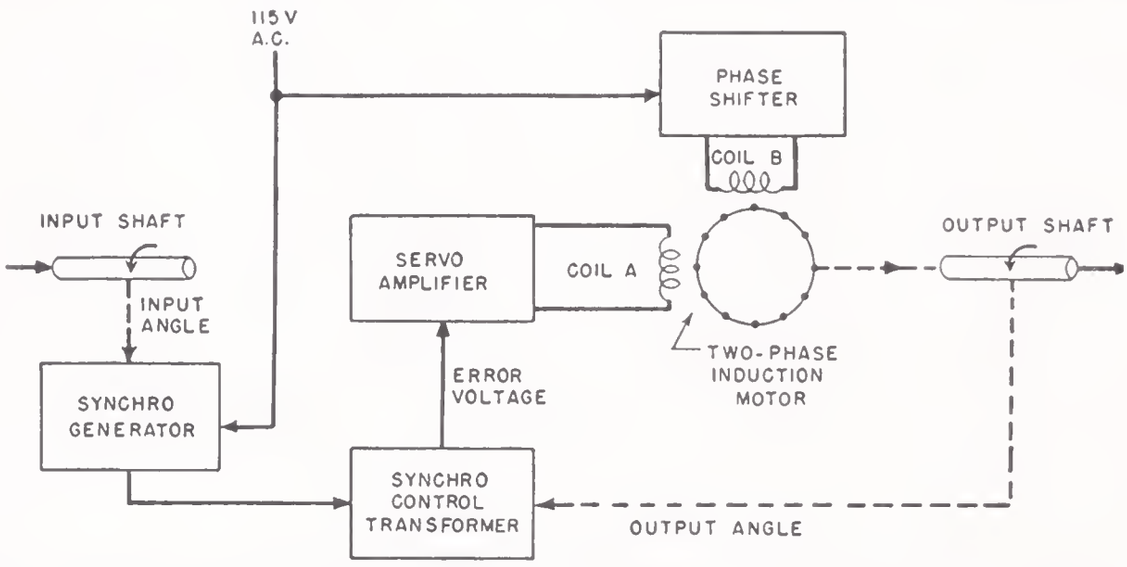


FIG. 43. Servo using a two-phase induction motor with controlled voltage applied to one stator winding.

is illustrated by the circuit of Fig. 43. The error voltage from the synchro control transformer is amplified and applied to coil A of the induction-motor stator, and a fixed alternating voltage 90 deg out of phase with the synchro voltages is applied to coil B.

If the error is zero, the magnetic flux in the induction motor is established by coil B alone and is an alternating flux the axis of which does not rotate. No torque is produced in the induction motor by such a flux. If an error exists, coil A carries current, and a rotating flux is produced

that drives the induction motor in the direction to correct the error. Reversing the direction of error reverses the phase of the current in coil A and thus reverses the direction of rotation of the motor.

The magnitude of the current in coil A, of course, depends upon the amount of the error. The equality of currents in coils A and B assumed in Art. 18 therefore does not usually exist. As a result, the field in the induction motor varies in magnitude as it rotates. For small error angles, the field has a maximum value at the instant of the cycle at which it is along the axis of coil B and a much smaller value when it is along the axis of coil A. Such a field is said to be elliptic, because it can be represented by a rotating vector the tip of which traces an ellipse as the vector

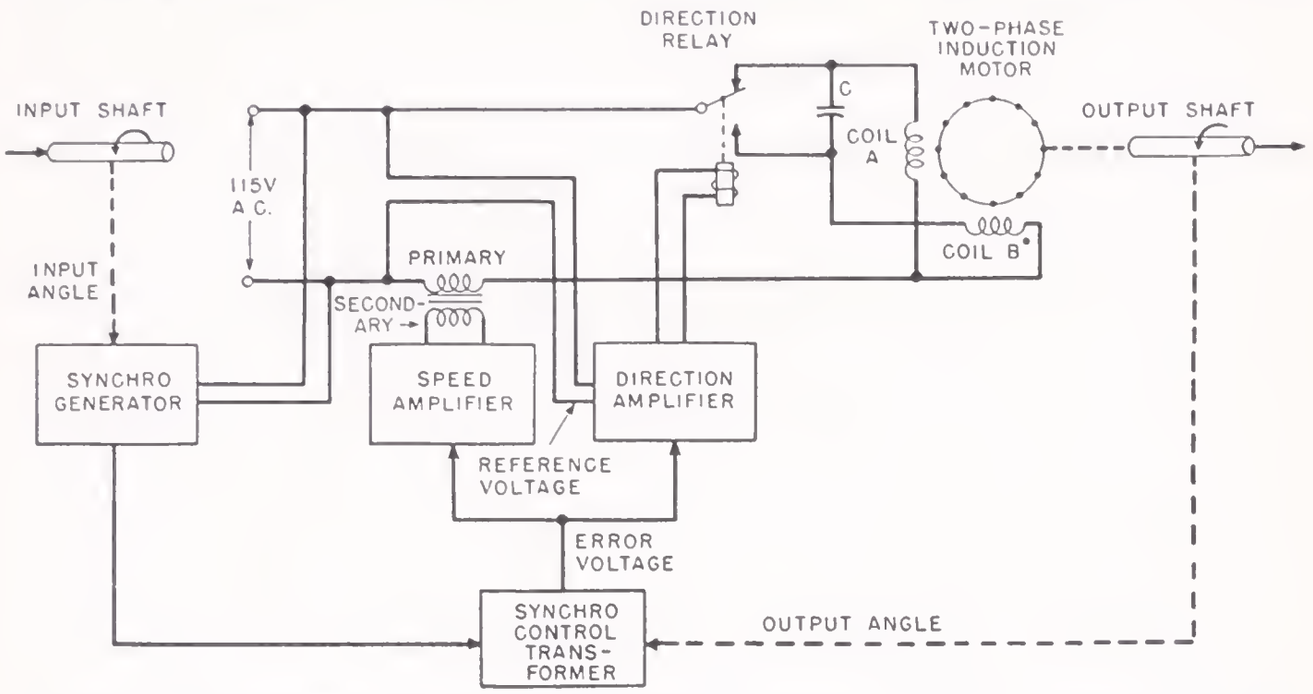


FIG. 44. Servomechanism using two-phase induction motor with speed amplifier and direction relay.

turns. If a motor with a high-resistance rotor is employed, the effectiveness of the elliptic field in producing torque is very nearly proportional to the minimum amplitude of the flux vector during its rotation. Therefore the torque of the servo motor is roughly proportional to the error angle, and smooth control of the output shaft is obtained.

Another method of controlling a two-phase induction motor is indicated in Fig. 44. The a-c power line is connected through the primary winding of a transformer to both stator coils of the induction motor, the 90-deg phase shift being provided by capacitor C. For one direction of error, the direction relay is unenergized, capacitor C is in series with coil B, and the current in coil B leads that in coil A. For reversed error, the direction relay is energized by current from the direction amplifier, the capacitor is connected in series with coil A, and the current in coil A leads that in coil B. Thus the direction of rotation of the motor is made to reverse when the direction of the error reverses.

The torque of the motor is controlled by varying the current that can pass through the transformer in series with the motor. The secondary voltage of the transformer is applied between plate and cathode of an amplifier tube in the speed amplifier. When the error is zero, this tube is biased beyond cutoff, and the secondary circuit of the transformer is open. Consequently very little primary current can flow. When an error in either direction exists, the bias voltage is changed and the tube is allowed to conduct. A motor current dependent upon the amount of the error is obtained, and the torque increases as the error increases.

20. Transient Oscillations in Servomechanisms. If the input shaft of a servomechanism is turned suddenly to a new position, the output shaft is often found to oscillate several times about its new position before

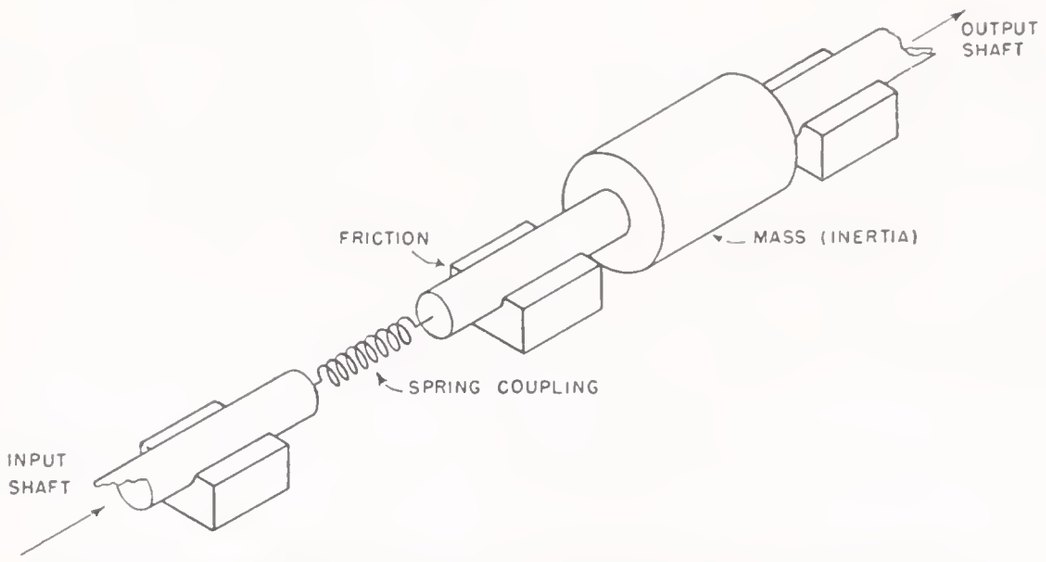


FIG. 45. Spring-coupled system analogous to a servomechanism.

coming to rest. Such behavior may be understood if the servo is compared with the mechanical system of Fig. 45.

The inertia and friction at the output shaft of the system in Fig. 45 correspond to the inertia and friction at the output shaft of a servo. The torque exerted by the spring on the output shaft in the figure is proportional to the error angle between output and input shafts and, therefore, is analogous to the torque supplied by a servo motor. An important difference between Fig. 45 and a servo is that in the figure the torque exerted by the spring on the output shaft equals the torque applied to the input shaft whereas torque amplification is obtained in a servo.

The output shaft in Fig. 45 follows motion imparted to the input shaft. If, however, the input motion is erratic or jerky, the output shaft may be expected to oscillate somewhat about the position of the input shaft. The frequency of the oscillations is determined by the stiffness of the spring and the inertia at the output shaft. The oscillations may be underdamped, critically damped, or overdamped, as in an *R-L-C* circuit (see Art. 18, Chap. II), depending upon the amount of the output-shaft

friction. Similarly, the oscillations obtained at the output shaft of a servomechanism have a frequency dependent upon the inertia of the servo motor and load and upon the stiffness of the controller. The amount of damping of the servo oscillations depends upon the friction at the output shaft.

An idea of the nature of the response of a servomechanism may be had from the curves of Fig. 46. The dashed-line plot of the input angle θ_i represents a step function or sudden displacement of the input shaft from one position to another. The solid-line plot of the output angle θ_o indicates the resulting motion of the output shaft, the various curves being drawn for different amounts of friction at the output shaft. If little friction is present, the oscillations begin with large amplitude and die away slowly, as in curve A. Increasing the friction causes the rate of

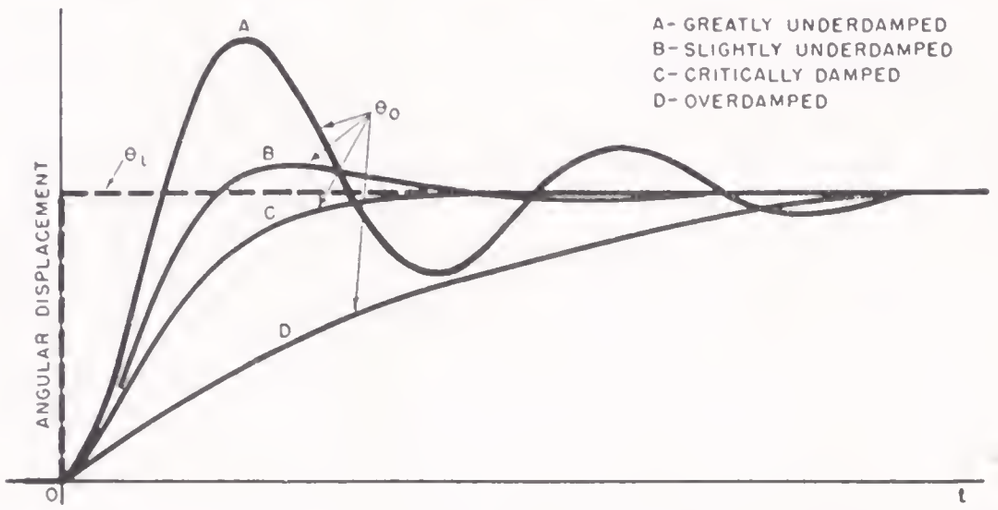
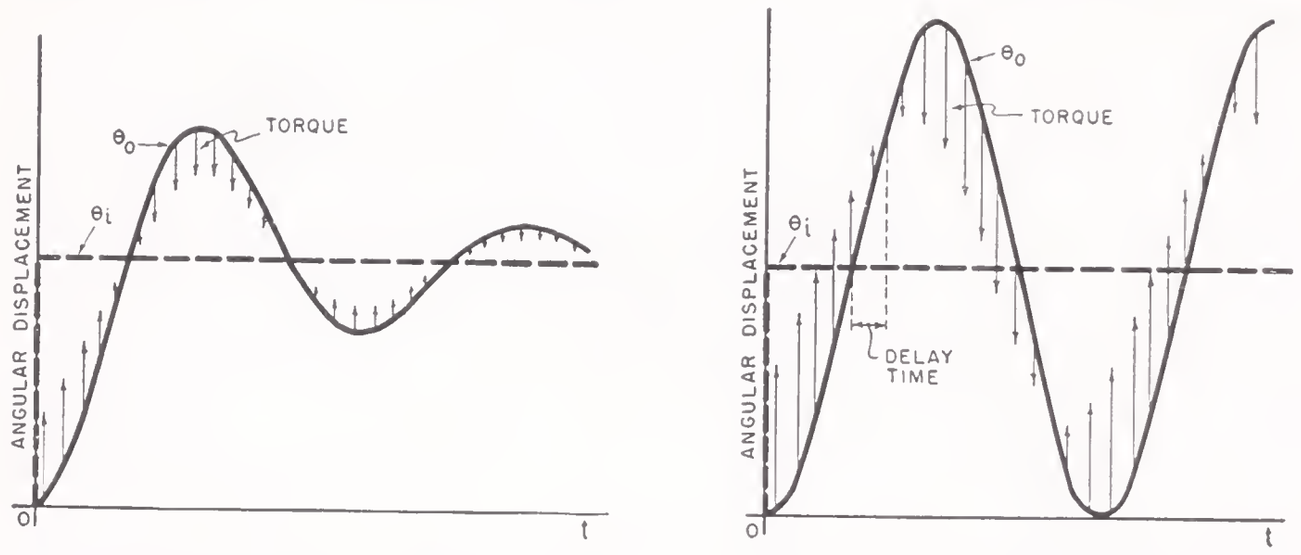


FIG. 46. Response of a servo to a sudden displacement of the input shaft.

decay of the oscillations to increase and the amount of overshoot to become less; also the frequency of oscillation decreases somewhat. Critical damping is obtained when the friction is barely sufficient to prevent oscillation. With greater amounts of friction, the output angle approaches its final value slowly and without oscillation.

21. Sustained Oscillations in Servomechanisms. The mechanical system in Fig. 45 is not completely analogous to a servomechanism. If a step function of angle is applied to the spring-mass system, oscillations of the output shaft may result, but the oscillations must always die away. During oscillation, energy is transferred back and forth between the kinetic energy of the rotating mass and the energy of the stressed spring, and some energy is lost because of friction. The oscillations must decrease in amplitude, because no source of energy is present. In a servomechanism, however, the controller is a power source and under certain conditions may supply the energy dissipated by friction forces. Thus oscillations of constant amplitude may be sustained. Oscillatory motion of a servomechanism is often referred to as *hunting*.

Whether or not a servomechanism can sustain oscillations depends upon whether or not a delay exists between the time that an error occurs and the time at which a torque is applied to correct the error. If the controller supplies a torque proportional to the *present* error, the forces acting on the output shaft are the same as those acting on the output shaft of Fig. 45, and oscillations cannot be sustained. The curves of Fig. 46 apply to such a system. Curve *A* from Fig. 46 is redrawn in Fig. 47*A* with arrows to indicate the direction in which the controller torque urges the output shaft. Note that the torque reverses when the error reverses and is at every instant in the correct direction to reduce the error. If the controller supplies a torque proportional to *past* error, as in Fig. 47*B*, the torque is in the direction to increase the error for a



(A) TORQUE PROPORTIONAL TO PRESENT ERROR (B) TORQUE PROPORTIONAL TO PAST ERROR

FIG. 47. Effect of delay time in controller.

short period after the zero-error position is crossed. This wrongly directed torque tends to increase the overshoot in each cycle of the oscillation, and if the delay is sufficient, it may cause the oscillations to grow to large amplitude instead of decaying.

The diagrams of Fig. 48 show more details of the relationships involved in a sustained oscillation. Assume that the input shaft is held stationary with $\theta_i = 0$ and that because of delay in the controller the output shaft is oscillating sinusoidally at constant amplitude. The output and error angles are then sinusoids, as are the output velocity and acceleration and the controller torque. Each of these quantities may thus be represented by complex numbers, and the diagrams of Fig. 48 may be used to show the phase and magnitude relationships. The output angle θ_o is taken as the reference sine wave and drawn in the zero-angle position in Fig. 48*A*. The error angle $\theta_e \cong \theta_i - \theta_o$ is equal to $-\theta_o$ and is drawn in the 180-deg position. The angular velocity ω_o of the output shaft is the derivative of θ_o. Since the derivative of $\sin 2\pi ft$ is $2\pi f \cos 2\pi ft$, ω_o leads θ_o

by 90 deg and has an amplitude proportional to the frequency, as indicated in Fig. 48*A*. Similarly, the output angular acceleration α_o (the derivative of ω_o) leads θ_o by 180 deg and is proportional in amplitude to the square of the frequency.

In order that oscillations may be sustained, the controller must supply the torque required by the friction and inertia at the output shaft. If viscous friction is assumed, the friction torque is at every instant proportional to the velocity of the output shaft and therefore is in phase with ω_o, as indicated by T_ω in Fig. 48*B*. The inertia torque is proportional at each instant to the acceleration of the output shaft and is therefore

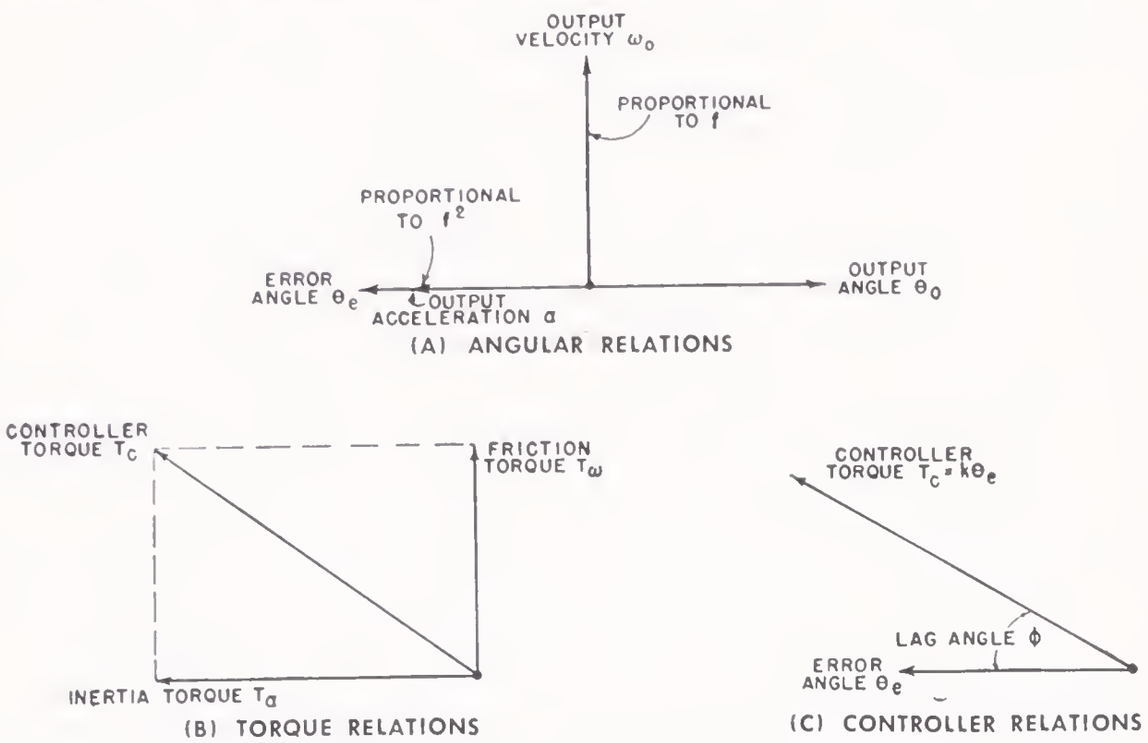


FIG. 48. Phase relations during sustained oscillations in a servomechanism.

represented by T_α in the figure. The resultant of T_ω and T_α is the torque T_c that the controller must supply.

The relationship of the controller torque to the error angle is shown in Fig. 48*C*. Because of friction torque, T_c must lag θ_e by the angle ϕ. Therefore, delay in the controller is necessary if sustained oscillations are to be possible.

The frequency of the sustained oscillations depends upon the inertia at the output shaft and upon the gain and delay time of the controller. The acceleration torque T_α is proportional to α_o, which in turn is proportional to the square of the oscillation frequency. Therefore, the oscillation frequency is determined by the requirement that T_α must equal the component of T_c along a horizontal axis in Fig. 48. This requirement may be used to show that the frequency of oscillation is increased (1) by increasing the gain of the amplifiers in the controller (which increases the constant k in Fig. 48*C*) or (2) by decreasing the inertia at the output shaft (which

decreases the amount of torque T_α obtained for a given output acceleration α_o). The frequency is also somewhat increased if the delay time of the controller is decreased, thereby decreasing the lag angle ϕ in Fig. 48*C*.

The possibility of sustained oscillations depends upon the friction at the output shaft and upon the gain and delay time of the controller. For sustained oscillations to exist, the component of T_c along a vertical axis in Fig. 48 must equal the friction torque T_ω. If this component of T_c is less than T_ω, then the energy supplied by the controller is less than that dissipated by friction, and oscillations die away. If the vertical-axis component of T_c is greater than T_ω, oscillations grow in amplitude until the controller output is reduced by overload to the value needed to sustain oscillations. Examination of Fig. 48 shows that sustained oscillations may be prevented by (1) increasing the friction at the output shaft, (2) decreasing the gain of amplifiers in the controller,[1] or (3) decreasing the delay time of the controller.

22. Antihunt Devices. Because there is always some delay in the controllers of servomechanisms, all servos will oscillate persistently if the gain of the amplifier in the controller is made sufficiently high. The highest possible gain is always desired, because high gain increases the speed of response by increasing the resonant frequency of the system and because high gain minimizes errors caused by load torques and miscellaneous effects, such as residual magnetism in an Amplidyne and unbalance of the two tubes in a phase-sensitive rectifier. Accordingly, special devices or networks are often included in servomechanisms which permit the gain to be increased without sustained oscillations occurring. Such damping or antihunt devices permit high gain and nearly critical damping to be realized simultaneously and therefore improve the performance of servomechanisms.

Mechanical Friction. Since sustained oscillations may be prevented by making the friction torque T_ω in Fig. 48 greater than the vertical-axis component of the controller torque T_c, an obvious method of stabilizing a servo system is to increase the mechanical friction at the output shaft. By sufficiently increasing this friction, not only can oscillations be prevented but also any desired degree of damping can be obtained.

For two reasons large amounts of mechanical friction are seldom used to damp servomechanisms. One reason is that such damping is very wasteful of power. The servo motor must produce many times the power required to turn the load—most of its output is converted into

[1] The torque T_ω does not remain constant when the factor k is changed. If k and therefore T_c are reduced by a factor of 4, for example, T_α is also reduced by 4 and the frequency of oscillation by $\sqrt{4} = 2$. Consequently, T_ω is reduced by a factor of 2. However, because the vertical-axis component of T_c varies in direct proportion to k while T_ω varies as $\sqrt{k}$, oscillation can be prevented by reduction of k.

heat by the damping friction. The other reason is that *velocity errors* are increased by the friction; that is, the friction increases the angle by which the output shaft lags the input shaft when both are turning at constant speed. The damping friction causes a torque at the output shaft, and an error is required to make the servo motor supply this torque.

To avoid the velocity error and to reduce the power consumed by friction, inertia dampers like those used on synchro motors (see Art. 5) are sometimes employed. These devices can be used only on small, low-power, servo systems, however, for the inertia of the damping device must be considerably larger than that of the servo motor and its load.

Electrical Damping of D-c Motor. Because of the voltage generated in the armature conductors of a d-c motor during rotation, electrical damping is obtained in servo systems that control the armature voltage of the motor. Consider, for example, the Amplidyne servo of Fig. 37, and assume that the servo-amplifier is disconnected from the Amplidyne so that no voltage is generated in that machine. If the motor armature is rotated, a voltage is generated in its conductors, and this voltage causes current to flow in the motor and Amplidyne armatures. The armature current causes a torque that opposes the rotation of the motor and is proportional in magnitude to the speed of rotation. The torque is related to the motion of the output shaft in the same way as is the torque of mechanical friction; thus an effect equivalent to mechanical damping is obtained electrically. When the servo loop is closed, two voltages in series send current through the Amplidyne and motor armatures. One, the voltage generated in the Amplidyne, causes errors to be corrected; the other, the voltage generated in the motor, causes damping.

To obtain maximum damping effect from the motor induced voltage, the armature resistance of the motor and the output resistance of the machine or amplifier that supplies power to the motor should be made as small as possible. Frequently, a negative-feedback circuit is employed to reduce the output resistance of the source of motor power.

A typical negative-feedback *antihunt circuit* is shown in Fig. 49 applied to an Amplidyne servomechanism. Assume, first, that capacitor C in the feedback circuit is short-circuited. The feedback is negative because connections are made so that the part of the Amplidyne voltage fed back opposes the effect of the error signal in unbalancing the currents of the tubes. If, for example, the error-signal voltage applied to V_1 is in phase with the plate supply voltage, so that the plate current of V_1 is greater than that of V_2, the Amplidyne voltage has the polarity marked on the figure and the feedback voltage tends to reduce i_{b1} and to increase i_{b2}. As in a cathode follower, the negative feedback makes the output voltage of the Amplidyne very nearly independent of the output current, which is equivalent to making the output resistance low. If a large current is

drawn from the Amplidyne, causing the output voltage to decrease a little, the feedback voltage is decreased and a greater difference of control-field currents is produced, which minimizes the decrease of output voltage. By means of this antihunt circuit, the Amplidyne output resistance may be reduced to such an extent that the damping currents are limited almost entirely by the armature resistance of the motor.

This form of electrical damping has an advantage over mechanical friction in that waste of the motor power is avoided. It has the same disadvantage as friction damping, however, in that a velocity error is produced. Thus in the circuit of Fig. 49 (with C short-circuited), if the input and output shafts are turning at constant speed, the Amplidyne must supply to the motor a direct voltage proportional to the speed. The

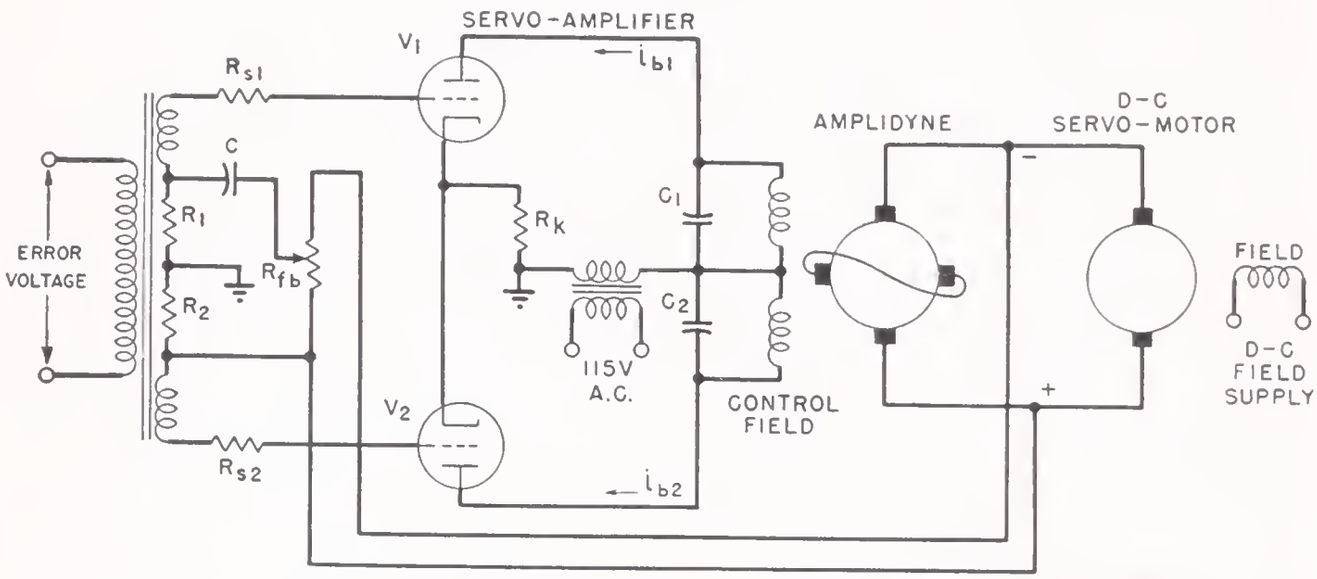


FIG. 49. Negative-feedback antihunt circuits for Amplidyne servomechanism.

part of this voltage fed back to the input of the servo-amplifier opposes the error signal and makes a greater error signal necessary to maintain an unbalance of currents in V_1 and V_2.

The velocity error may be avoided by use of capacitor C, as in Fig. 49. This capacitor and the resistors R_1 and R_2 are large, so that at the frequency for which oscillation might occur the antihunt voltage is fed back almost as though the capacitor were short-circuited. In a steady-state motion, however, the Amplidyne voltage is constant and is blocked by the capacitor. The antihunt circuit is analogous to the inertia damper in that damping forces are produced when they are needed to prevent oscillation and are removed in constant-speed rotation to prevent velocity errors.

In a servo using an antihunt circuit of this kind, the gain of the servo-amplifier is usually made high enough to cause sustained oscillations when no antihunt voltage is used. The oscillations can be stopped by applying the antihunt voltage, that is, in the circuit of Fig. 49, by moving the tap

on R_{fb} upward from the bottom. If, however, the tap is raised too far, a new kind of oscillation at a much higher frequency than the hunting is often obtained. This type of oscillation results because the servo-amplifier, Amplidyne, and antihunt circuit form a closed feedback loop. This closed circuit, like any other negative-feedback amplifier, will oscillate if the feedback is made sufficiently great. At some relatively high frequency, sufficient phase shift occurs in the servo-amplifier and Amplidyne to cause the feedback to be positive.

Lead Networks and Derivative Control. The mechanical-friction and electrical-damping antihunt systems increase the damping forces applied

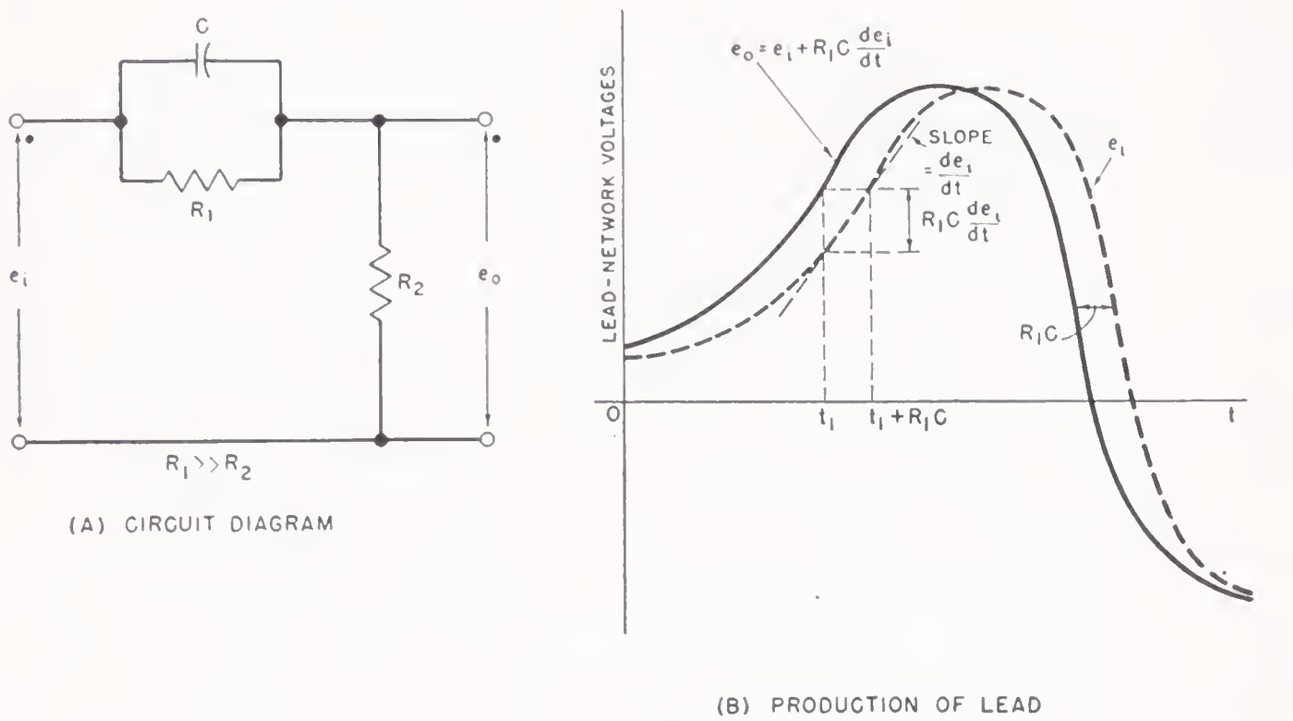


FIG. 50. D-c lead network.

to the output shaft in order to make possible a high gain of the servo-amplifier without excessive tendency of the servo to oscillate. An alternate method of obtaining high gain without oscillation is to decrease the delay in the controller. To this end controller components having the smallest possible inherent delays are chosen, and sometimes *lead networks* are inserted in the controller. A lead network produces a leading phase shift of the error signal within the band of frequencies in which oscillation is possible. Thus the lead network compensates for the phase lag in other parts of the controller.

The circuit diagram of a simple lead network is drawn in Fig. 50*A*. Resistor R_2 is assumed to be very small so that the voltage across R_1 and C can be taken to equal e_i. Thus the current through resistor R_1,

$$i_{R1} \approx \frac{e_1}{R_1}$$

is proportional to the input voltage, and the current through C,

$$i_C \approx C \frac{de_i}{dt}$$

is proportional to the derivative of the input voltage. The output voltage e_o is the voltage produced across R_2 by the sum of i_{R1} and i_C,

$$e_o = R_2 \left(\frac{e_i}{R_1} + C \frac{de_i}{dt} \right) = \frac{R_2}{R_1} \left(e_1 + R_1 C \frac{de_i}{dt} \right) \tag{5}$$

and contains two parts, one proportional to the input voltage and the other proportional to the derivative of the input voltage. A servo in which the error voltage is sent through a lead network is therefore said to have *proportional plus derivative control*, or *error plus error-rate control.* Sometimes, instead of a lead network, a device sensitive to the derivative of the error is included in a servo, and the derivative output added to the proportional output of the error detector. The total control signal is then of the same nature as the signal obtained by use of a lead network.

According to Eq. 5, the lead network has the effect of attenuating the input signal greatly (R_2/R_1 is much less than 1) and of adding the term $R_1C\ de_i/dt$ to e_i. This added term advances the wave of e_i along the time axis by the time interval R_1C, as illustrated in Fig. 50*B*, provided the e_i wave does not contain very abrupt changes. In the figure, an e_i wave is assumed for which the derivative de_i/dt does not change appreciably in any time interval of duration R_1C. For such a wave, the quantity $R_1C\ de_i/dt$ of Eq. 5 is very nearly equal to the change of e_i in the interval R_1C. Thus at the instant t_1, the voltage $e_i + R_1C\ de_i/dt$ has the same value as e_i has at the instant $t_1 + R_1C$. The condition that the derivative of the e_i wave must not change too abruptly is the same as the requirement that the wave must not contain Fourier components of too high frequency. If R_1C is sufficiently short, the band of frequencies in which the lead is produced is wide enough to include the frequency of possible oscillation, and the lead network can compensate for lag in other parts of the servo.

The lead network of Fig. 50*A* is called a *d-c lead network*, because it operates upon a direct voltage, the magnitude and sign of which depend upon the error. A different network is needed in servomechanisms like that of Fig. 43 in which a two-phase induction motor is used. Such servos require *a-c lead networks*, that is, networks which operate upon an a-c wave modulated by the error signal. A possible error voltage from a control transformer is drawn in Fig. 51; it is a power-frequency wave, the magnitude and phase of which indicate the amount and direction of the error. The desired output of an a-c lead network to which this error

voltage is applied is also drawn in Fig. 51. Note that the phase of the power-frequency wave remains unaltered (so that the phase relations of currents in the induction motor are correct for producing a rotating field) but the *modulation envelope* representing the error is shifted ahead by the lead time.

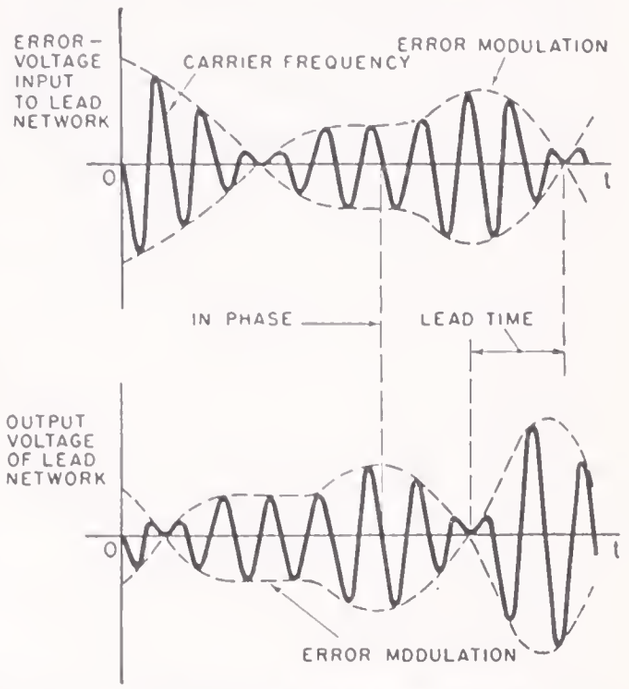


FIG. 51. Effect of an a-c lead network on an a-c error signal.

The characteristics that an a-c lead network must have to produce such an effect may be seen if the error voltage is separated into its sinusoidal components. Suppose that the error itself varies sinusoidally at a low frequency. The amplitude of the a-c error signal may be represented by $E_m \cos \omega t$, where ω is the low angular frequency of the error variation. If ω_o is the angular frequency of the synchro supply voltage, the control-transformer voltage applied to the lead network is

$$e_i = (E_m \cos \omega t) \cos \omega_o t$$

or, by use of a trigonometric identity,

$$e_i = \frac{E_m}{2} \cos (\omega_o + \omega)t + \frac{E_m}{2} \cos (\omega_o - \omega)t \tag{6}$$

Equation 6 shows that the error signal may be considered as the sum of the two side-band waves, the frequency of one being above the line frequency by an amount equal to the error frequency and the frequency of the other being below by the same amount. This wave is like the amplitude-modulated waves used in radio and radar (see Art. 4, Chap. VI), the power frequency corresponding to the carrier frequency and the error frequency corresponding to the modulating frequency. There are two important differences, however, between servo error signals and modulated radio waves: (1) The radio wave is the sum of a carrier wave and side-band waves; the error signal comprises only the side bands. For this reason the error signal is called a suppressed-carrier amplitude-modulated wave. (2) The actual values of frequencies in the error signal are tremendously lower than those in radio waves. For example, the error signal might be a 60-cps wave modulated by error frequencies of less than 10 cps.

An r-f pulse is shown in Art. 4 of Chap. VI to have a frequency spectrum like the spectrum of the video pulse forming its envelope, except that frequencies in the r-f pulse are measured up and down from the carrier frequency instead of up from zero. The spectrum of any modulated wave is related to the spectrum of its envelope in the same manner. Suppose an a-c error signal is impressed on an a-c lead network and a d-c signal representing the same error is impressed on a d-c lead network, both networks having the same lead time. The frequency spectra of the two input waves are the same except for the shift due to modulation, and

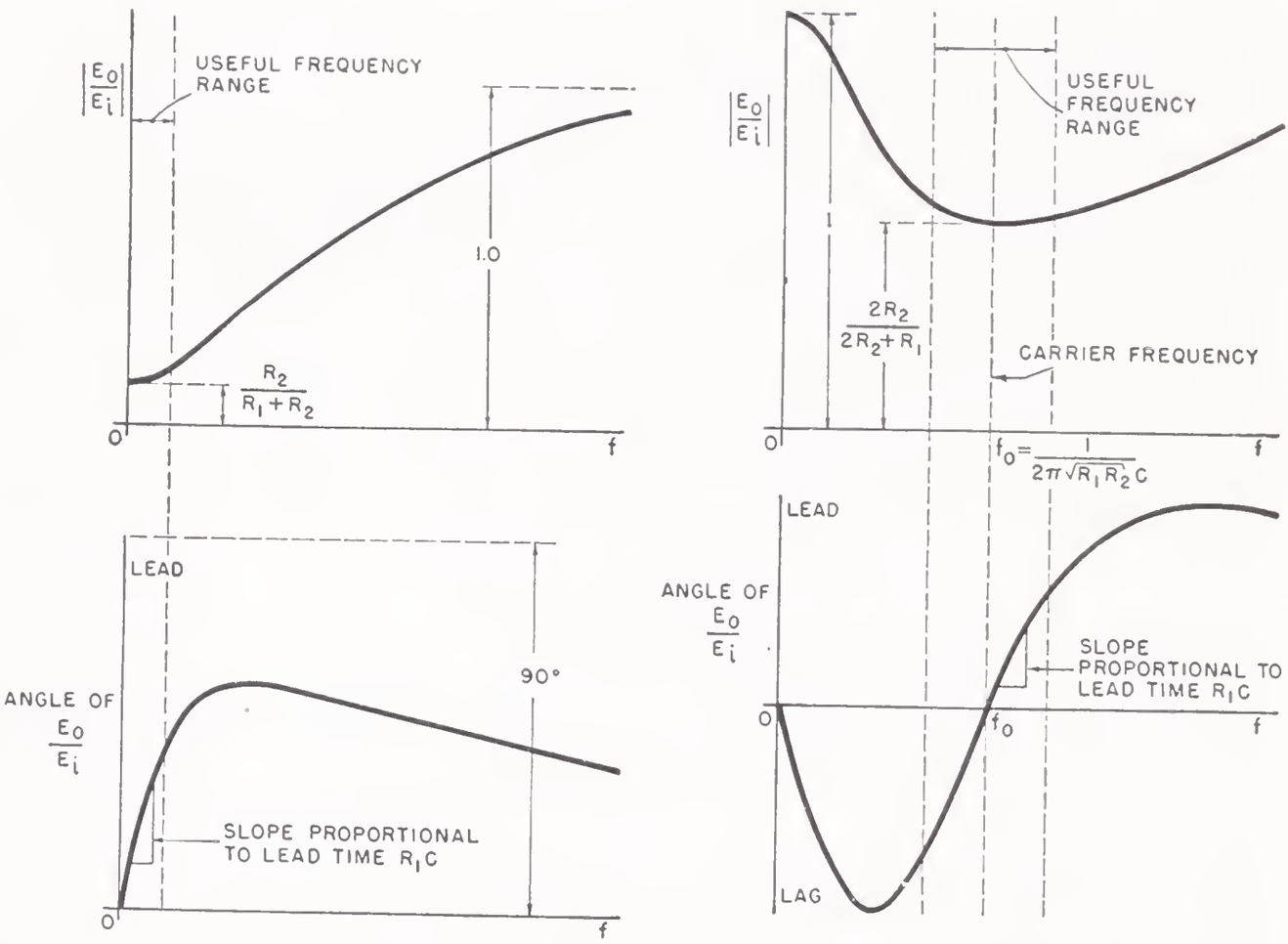


FIG. 52. Frequency characteristics of lead networks.

the same relation is true for the spectra of the two output waves. Accordingly, the frequency-response curve of an a-c lead network must be like that of a d-c lead network, except that modulation frequencies should be measured up and down from the carrier frequency, instead of up from zero. Figure 52*A* shows the amplitude and phase characteristic of the d-c lead network of Fig. 50, and Fig. 52*B* shows the corresponding characteristic of an a-c lead network. The minimum point in the amplitude-response curve occurs at the line frequency f_o.

An *R-C* network having the frequency response of Fig. 52*B* can be obtained by combining an *R-C* low-pass network with an *R-C* high-pass network. The low-pass network of Fig. 53*A* has a frequency character-

istic (Fig. 53*B*) like the low-frequency portion of the curve in Fig. 52*B*. The high-pass network of Fig. 54*A* has a frequency characteristic (Fig. 54*B*) like the high-frequency portion of the curves in Fig. 52*B*. If the two networks are combined as in Fig. 55*A*, the frequency response obtained is that of Fig. 52*B*. At low frequencies R_2 may be considered to be a short circuit, as its resistance is small compared with the reactances of C_1 and C_2; thus the network reduces to that of Fig. 53*A*. At high frequencies R_1 may be considered an open circuit, because its resistance is large compared with the capacitor reactances; thus the network reduces to that of Fig. 53*B*. In a-c lead networks C_1 is usually made equal to C_2, and the network is usually drawn as a bridged T, as in Fig. 55*B*.

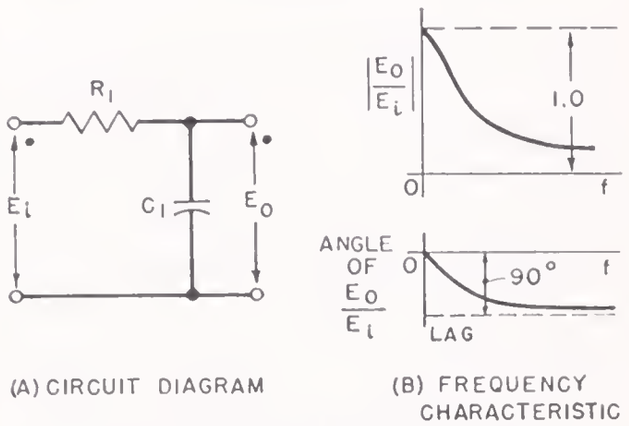


FIG. 53. Low-pass *R-C* network.

The relationship of the network parameters to the line frequency and to the lead time are indicated in Figs. 52*B* and 55*B*. Note that the slope of the phase-response curve is a measure of the lead time; in fact, this slope equals the lead time if the curve is a plot of phase shift in radians against angular frequency in radians per second. Note also that, to produce a large lead time, the dip in the amplitude-response curve must be deep and have steeply sloping sides. Circuits with amplitude-response curves having a dip at the line frequency produce a lead of the modulation envelope; resonant circuits that produce an increase of response at the line frequency cause a delay of the modulation envelope.

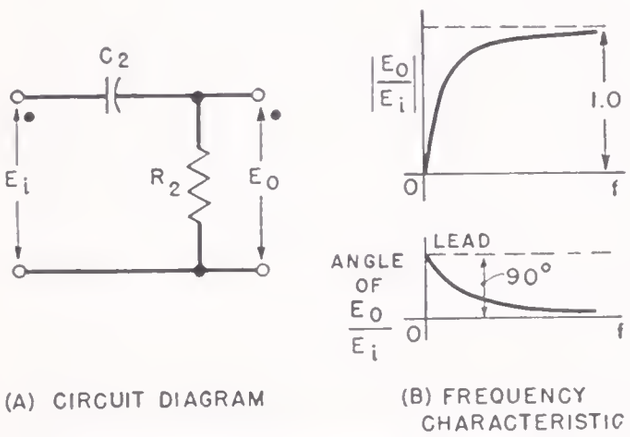


FIG. 54. High-pass *R-C* network.

Because a dip in the amplitude-response curve is associated with the lead time, there is a practical limit to the amount of servo damping that can be obtained with a lead network. If the lead time is made too great, small extraneous voltages of frequency far removed from the line frequency are amplified to such an extent that they interfere with the operation of the servo. The voltages obtained from synchros are not perfect sine waves; the extraneous voltages applied to a-c lead networks are mainly the harmonic-frequency components of these waves. If d-c lead networks are used following a phase-sensitive rectifier in a servomecha-

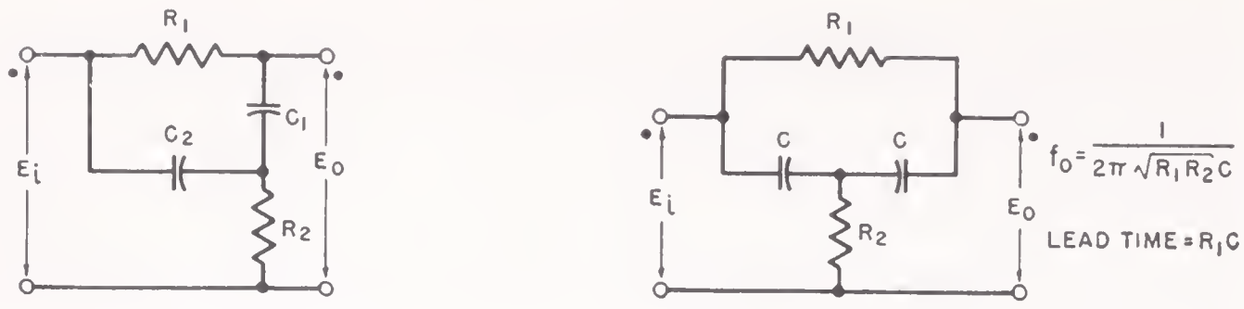


FIG. 55. A-c lead network.

nism, the ripple voltage of the rectifier is an extraneous voltage that is exaggerated by the network. An additional limit on the usefulness of an a-c lead network exists if the line frequency is not perfectly constant. Operation of the system may then be disturbed by the frequency shifting away from the midpoint of the dip.

A lead network has the same damping effect on servomechanism response as has the antihunt circuit of Fig. 49, p. 330. The two circuits differ, however, in their effect upon the errors of the servomechanism. The antihunt circuit always increases the errors. Without a blocking capacitor, it produces a velocity error; with the capacitor, an error occurs when the velocity is suddenly changed and persists until the capacitor has time to charge to a new voltage. The lead network, on the other hand, has no effect on steady-state errors and reduces transient errors. If a motion of the input shaft is suddenly started, and if the output does not follow immediately, the error begins to increase from zero. The resulting rate of increase of error causes a torque in the direction to correct the error, because of the derivative component of the lead-network output. This torque is in addition to the torque resulting from the proportional component of the lead-network output.

23. Applications of Servomechanisms. Servos in radar sets are used principally to control the motion of the antenna and to maintain synchronism between the antenna and the rotating coil of a plan-position indicator. If an antenna is required merely to rotate at constant speed, there is no need for a servo drive. If, however, several different motions are required, such as constant-speed rotation, scanning back and forth over a small sector, and tracking of a moving object, then a servo drive is useful. Deflection coils are light enough to be turned by synchros, and synchro generator-motor drives are often used with them (see Art. 12). However, greater accuracy can be obtained from servos, and deflection coils are frequently driven by low-power servos, usually employing small two-phase induction motors.

There are two general methods of using servos to drive antennas and deflection coils. In one, a large servo controls the antenna position and a

small servo (or a synchro system) maintains the coil in correspondence with the antenna position. The large servo can cause the antenna to follow arbitrary movements of a hand control, or the servo motor can be set to drive the antenna through specified scanning patterns automatically. In the second method, a single servo is used to move the antenna in correspondence with the deflection coil. The coil is then turned by a hand control, driven at constant speed, or made to sector scan by means of a small motor.

Special servos are needed to control the antenna of an automatic-tracking set. There is no input shaft; instead the antenna must follow the movements of a distant object that is being tracked. The antenna must move in both azimuth and elevation, and thus two servos are required, one to turn the antenna about each axis. Instead of a synchro, the entire radar set is the error detector, and conical scan is used to provide error signals for both servos. The modulation envelope of the series of echo pulses received from the object tracked is the error signal. If the object lies exactly on the axis of the cone of scan, all pulses have the same amplitude and the error signal is zero. If the object lies off this axis, an approximately sinusoidal variation of pulse amplitude is produced at a frequency equal to the rate at which the antenna beam traces its cone. Rectification of the series of pulses, therefore, produces an approximate sine wave of low frequency. The phase of this wave, relative to a voltage obtained from the device that causes the conical scan, indicates the direction from the cone axis to the object tracked. Therefore the wave may be separated into two components in quadrature, and the components used as individual error signals for the azimuth and elevation servos.

In more elaborate shipboard radar sets, servos are used to stabilize the antenna against motion of the ship. Suppose that the antenna is intended to scan the horizon. Rolling and pitching of the ship cause the beam to be directed above and below the horizon if no stabilization is provided. By means of servos controlled by gyroscopes, the antenna platform can be maintained level despite movements of the ship. Thus in surface search, the need for a broad beam in elevation may be avoided.

CHAPTER VI

RECEIVERS

The radio-frequency echo pulses returned by a distant object are similar to the transmitted pulses, but very much smaller—perhaps a billion billion times smaller in power. These tiny pulses are greatly amplified and are transformed into video pulses by the receiver. A voltage amplification of the order of 10,000,000 is required to produce an output video pulse of sufficient amplitude to deflect or intensify the electron beam of a cathode-ray tube. The receiver must accomplish this amplification with the least possible introduction of noise voltages.

A. GENERAL PRINCIPLES

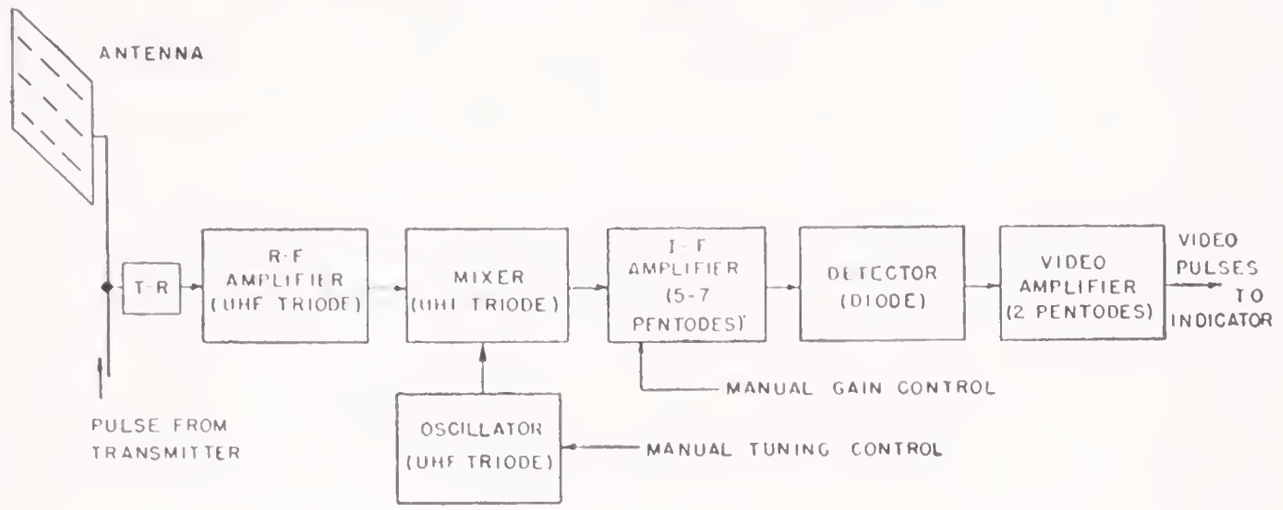
In preparation for the detailed description of receiver components, a discussion is given in this section of the over-all arrangement of the components and of some general principles of receiver analysis. These general principles include the application of Fourier analysis and complex-number calculations to pulse amplifiers, the equivalent circuits for tubes at receiver frequencies, and the effects of noise in a receiver.

1. Block Diagram. The *superheterodyne receiver* is used almost exclusively in radar systems, and it alone is discussed in this chapter. Other circuits (for example, superregenerative receivers and crystal-video receivers) are useful in special applications, but in normal radar service the superheterodyne has the important advantages of low noise, efficient amplification, reliable operation, and freedom from blocking (see Arts. 19 and 22).

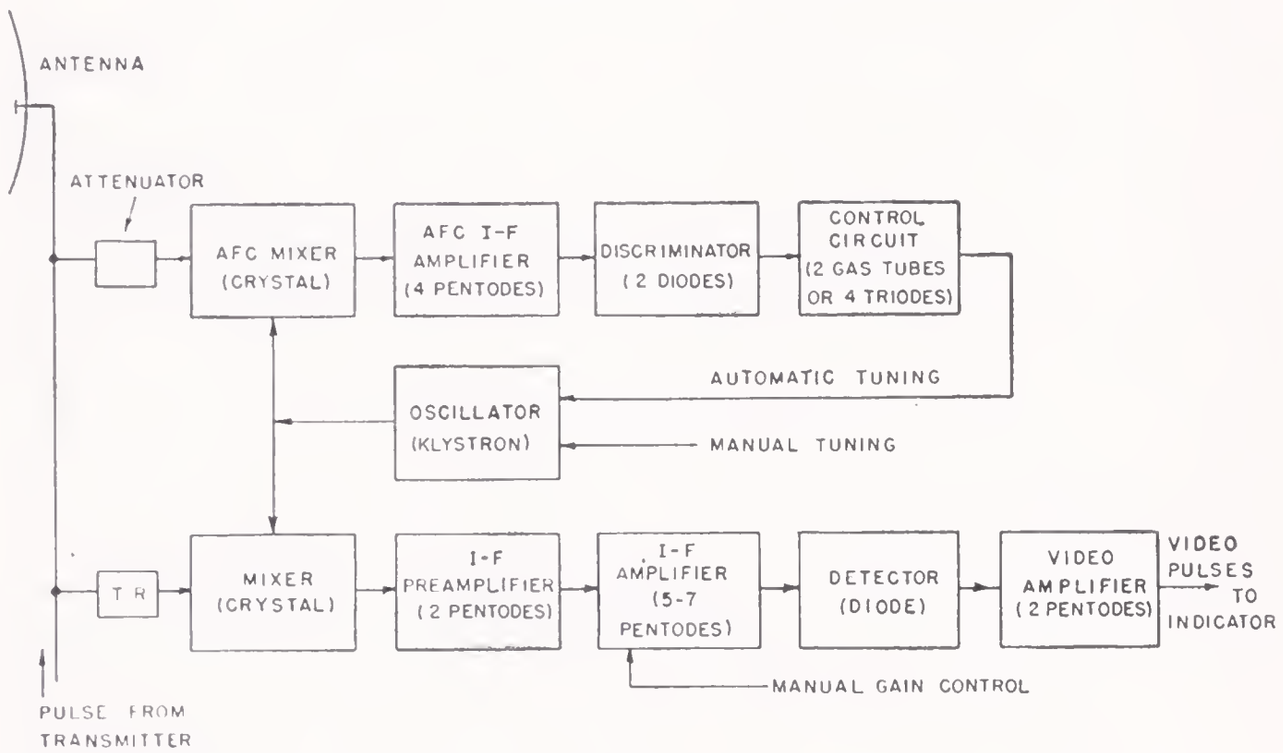
In a superheterodyne receiver, a continuous sine wave from a radio-frequency oscillator (see Fig. 1) is rectified by a nonlinear element in the mixer stage, and a steady direct current is produced. When signals are received, the tiny echo pulses are also applied to the mixer, and the amplitude of the sum of echo-pulse and oscillator waves varies over a small range as the two waves, which differ slightly in frequency, swing into and out of phase. If, for example, the input frequencies are 1,000 and 1,030 Meps, a 30-Meps ripple appears in the mixer output whenever an echo pulse is present. This ripple constitutes an intermediate-frequency pulse and is amplified by the i-f stages. Most of the amplification of the receiver is provided by these stages, and from them an i-f pulse of roughly a volt in amplitude is obtained. The detector is a

second rectifying element and produces a d-c output whenever an i-f pulse is applied to it. Thus the video pulses are formed.

The detailed arrangement of components varies from one receiver to another, but for radar frequencies below about 1,000 Mcps, receivers



(A) TYPICAL UHF RECEIVER



(B) TYPICAL MICROWAVE RECEIVER

FIG. 1. Block diagrams of radar receivers.

often follow the general plan of Fig. 1*A*. For higher radar frequencies, the diagram of Fig. 1*B* is typical.

The r-f amplifier, mixer, and oscillator in Fig. 1*A* use special tubes designed for ultrahigh-frequency operation. The echo pulse is amplified before being applied to the mixer for only one reason: Available tubes produce less noise in amplifier operation than in mixer operation. Con-

sequently, a smaller echo pulse can be detected if it is amplified before it enters the noisy mixer stage. Because echo pulses of only one frequency are to be received, the r-f amplifier and mixer circuits are fixed tuned—screw-driver adjustments suffice for alignment purposes. Small changes of transmitter frequency that occur in operation are compensated by adjustment of the local-oscillator frequency. The gain control usually varies the grid-bias voltage or screen-grid supply voltage for two or more stages of the i-f amplifier.

No r-f amplifier is drawn in Fig. 1*B* because available tubes introduce excessive noise when operated in the microwave bands. The echo pulse is applied directly to a silicon-crystal mixer, which introduces very little noise. The mixer does, however, attenuate the signal and therefore exaggerates the importance of noise in the i-f amplifier. To minimize the signal attenuation, the mixer and the first stage or two of the i-f amplifier (the preamplifier) are usually placed very close to the t-r device. A disadvantage of the crystal mixer is that the t-r device must block transmitter power with extreme effectiveness to prevent damage to the delicate crystal.

Receiver bandwidths do not increase in proportion to the radio frequency when the radio frequency is increased. Therefore, small percentage changes of transmitter or local-oscillator frequency have a greater detuning effect in higher frequency sets. To reduce tuning difficulties, therefore, automatic-frequency-control circuits are usually provided in microwave receivers, as shown in Fig. 1*B*. The oscillator, which is a klystron, can be tuned through a small range by adjustment of the voltage applied to the reflector electrode (see Art. 4, Chap. XI). A small portion of the transmitted pulse is combined with the local-oscillator voltage in the afc mixer to produce i-f pulses of the same carrier frequency as those in the main receiver channel. If this frequency is correct, the discriminator output is zero. Otherwise, the discriminator produces positive or negative video pulses, depending upon the direction of the error. These pulses actuate the control circuit and cause a change of reflector voltage in the direction to correct the frequency error.

The r-f amplifier, mixers, and local oscillators of Fig. 1 operate at frequencies of 200 to 30,000 Mcps and employ transmission-line sections and resonant cavities as tuning elements. The characteristics of these elements are explained in Chap. IX, and discussion of r-f receiver components is best postponed until after this explanation. The present chapter considers the low-frequency aspects of mixer operation and the operation of all subsequent portions of the receiver. The uhf oscillator of Fig. 1*A* is discussed in Chap. X, the klystron of Fig. 1*B* is described in Chap. XI, and the r-f amplifier of Fig. 1*A* and the high-frequency aspects of mixers are treated in Chap. XII.

2. Sine-wave Analysis. Signal voltages in the receivers are groups of more or less rectangular pulses representing echoes from objects along the path of the antenna beam. In preceding chapters, the response of a circuit to pulses is computed directly in terms of transients in the circuit. Such computations become exceedingly difficult in receiver circuits because of the many stages of a receiver. Calculation of the successive modifications of an originally rectangular pulse as it passes through the receiver would indeed be laborious. Furthermore, even so difficult a computation would yield only an incomplete picture of receiver performance, because actual echo pulses have a variety of different wave shapes, and to the echo signals must be added noise voltages of extremely erratic variation.

To avoid these difficulties, an entirely different procedure is used in the study of receivers. The response to unmodulated sine waves of various frequencies is determined, and the response to waveforms actually present is inferred from the sine-wave response. The advantage of this method is the relative ease with which sine-wave response can be measured and calculated. The correctness of the method depends upon two principles: (1) Any signal waveform in a receiver can be expressed as the sum of a group of sine waves. (2) The response of a *linear* circuit to the sum of several excitations is, according to the *principle of superposition*, the sum of the responses to the component excitations acting separately.

Because the superposition principle applies only to linear circuits, the method of determining pulse response by adding responses to sine-wave components cannot be applied to most of the circuits in the preceding chapters. Tubes in these circuits operate over a wide range of the plate characteristics and must be considered nonlinear devices. Sine-wave analysis is particularly adapted to the study of receivers actuated by *small-amplitude signals*. With such signals applied, most of the receiver tubes operate in small, nearly linear ranges of their characteristics and may be represented by linear equivalent circuits. Except for certain properties (see Art. 26) of the detector, the receiver is then equivalent to a network of linear elements.

The resolution of arbitrary signals into sinusoidal components is called *Fourier analysis*. If the signal resolved is a periodic wave, the frequencies of the sine waves obtained are multiples of the repetition frequency of the signal wave, and the result of the resolution is a *Fourier series*. If a periodic voltage e has a frequency f_r and an angular frequency $\omega_r = 2\pi f_r$, its Fourier series is

$$e = E_o + E_{m1} \sin (\omega_r t + \phi_1) + E_{m2} \sin (2\omega_r t + \phi_2) + E_{m3} \sin (3\omega_r t + \phi_3) + \cdots \quad (1)$$

The constant E_o is the *direct component* or average value of the wave, given by

$$E_o = \frac{1}{2\pi} \int_{-\pi}^{\pi} e\, d(\omega t) \tag{2}$$

and the sine terms are the *harmonic components*, or *alternating components*, of the wave. The first harmonic component, called the *fundamental*, has a frequency equal to that of e, a peak value E_{m1}, and a phase angle ϕ_1. The second harmonic has a frequency twice that of e, an amplitude E_{m2}, a phase angle ϕ_2, and so on for the other components.

In an alternate form of the Fourier series, each harmonic component of Eq. 1 is expressed as the sum of two sinusoidal waves in quadrature. For the fundamental, for example, the trigonometric expansion of the sine of the sum of two angles yields

$$E_{m1} \sin (\omega_r t + \phi_1) = (E_{m1} \cos \phi_1) \sin \omega_r t + (E_{m1} \sin \phi_1) \cos \omega_r t \tag{3}$$

and the fundamental is expressed as the sum of a sine wave of peak value $E_{m1} \cos \phi_1$ and a cosine wave of peak value $E_{m1} \sin \phi_1$. Similar resolution of all the harmonic components in Eq. 1 converts the Fourier series into the sum of a direct component, a series of sine terms, and a series of cosine terms.

The peak values of the sine and cosine terms of the series can be shown[1] to be

$$\left.\begin{array}{l}\text{Peak value of } k\text{th} \\ \text{sine component}\end{array}\right\} = E_{mk} \cos \phi_k$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} e \sin k\omega_r t\, d(\omega_r t) \tag{4}$$

$$\left.\begin{array}{l}\text{Peak value of } k\text{th} \\ \text{cosine component}\end{array}\right\} = E_{mk} \sin \phi_k$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} e \cos k\omega_r t\, d(\omega_r t) \tag{5}$$

In Eqs. 4 and 5, k may be assigned any integral value; therefore, these equations may be used to determine the coefficients of all the sine and cosine terms in the Fourier series. Once these coefficients are known, the peak values E_{mk} and phase angles ϕ_k in Eq. 1 may be computed from the identities

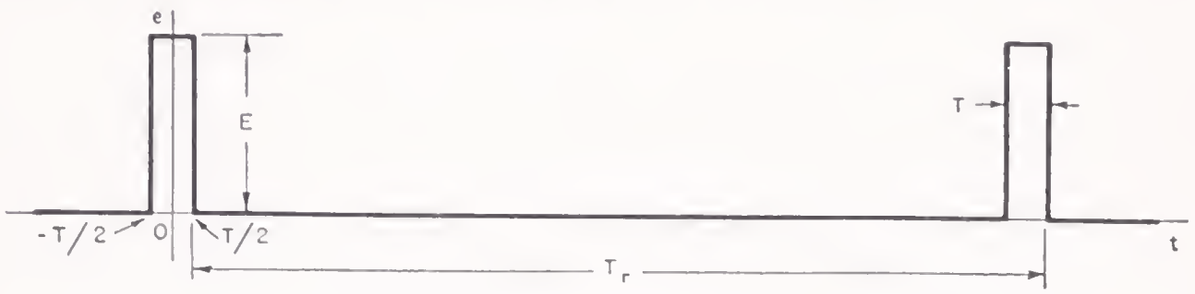
$$E_{mk} = \sqrt{(E_{mk} \sin \phi_k)^2 + (E_{mk} \cos \phi_k)^2}$$

and

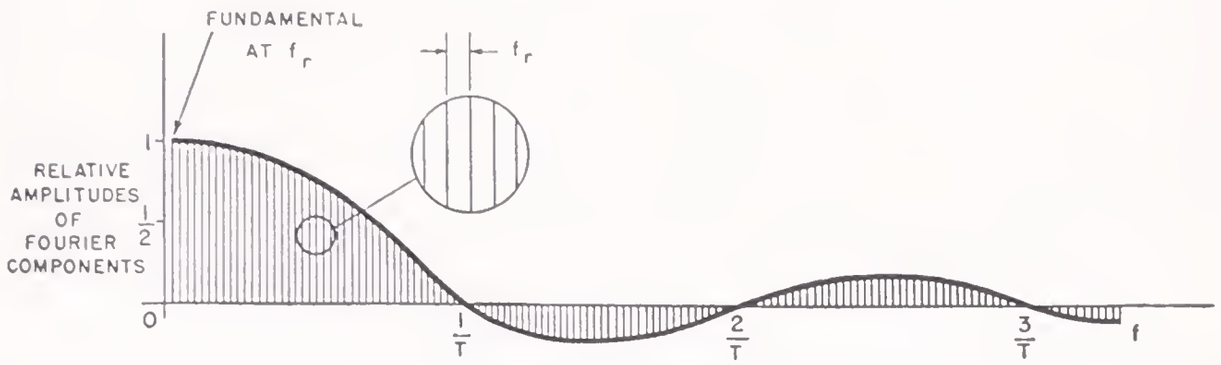
$$\phi_k = \tan^{-1} \frac{E_{mk} \sin \phi_k}{E_{mk} \cos \phi_k}$$

[1] Derivations of Eqs. 1, 2, 4, and 5 may be found in many mathematical texts, for example, P. FRANKLIN, *Differential Equations for Electrical Engineers*, 61–77 (John Wiley & Sons, Inc., New York, 1933).

3. Sine-wave Components of Video Pulses. A useful example of the application of Fourier series is the analysis of a video pulse (Fig. 2*A*) of duration T, repetition period T_r, and amplitude E. For convenience in the calculation of the Fourier-series components, the origin of the time scale is located at the center of the pulse. The pulse begins and ends at the instants of time $-T/2$ and $T/2$, or because the time interval T_r corresponds to an interval 2π of the angle $\omega_r t$, at the values $-(T/T_r)\pi$ and



(A) VIDEO-PULSE WAVEFORM



(B) FREQUENCY SPECTRUM

FIG. 2. Fourier components of video pulse.

$(T/T_r)\pi$ of $\omega_r t$. Thus the average value of the pulse waveform, according to Eq. 2, is

$$E_o = \frac{1}{2\pi}\int_{-(T/T_r)\pi}^{(T/T_r)\pi} E\,d(\omega_r t) = \frac{T}{T_r}E$$

Because of the location chosen for the origin of the time scale, the integration of Eq. 4 yields a value zero; thus the Fourier series contains only cosine terms. According to Eq. 5, the cosine coefficients are

$$\begin{aligned} E_{mk}\sin\phi_k &= \frac{1}{\pi}\int_{-(T/T_r)\pi}^{(T/T_r)\pi} E\cos k\omega_r t\,d(\omega_r t) \\ &= 2E\,\frac{T}{T_r}\,\frac{\sin k\pi T/T_r}{k\pi T/T_r} \end{aligned}$$

The Fourier series for the pulse of Fig. 2*A* therefore takes the form

$$e = 2E\,\frac{T}{T_r}\left(\frac{1}{2} + \frac{\sin \pi T/T_r}{\pi T/T_r}\cos\omega_r t + \frac{\sin 2\pi T/T_r}{2\pi T/T_r}\cos 2\omega_r t + \cdots\right) \quad (6)$$

A graphical presentation, called a *frequency spectrum*, of the frequencies and relative amplitudes of the components of the Fourier series is made in Fig. 2*B*. Vertical lines are drawn at positions along the frequency scale corresponding to the frequencies of the harmonic components, and the lengths of the lines are made proportional to the amplitudes of the components. Thus, at $f = 0$, a line one-half unit long is drawn representing the direct component. At $f = f_r$, a line representing the fundamental is drawn, and its length is $(\sin \pi T/T_r)/(\pi T/T_r)$, which is very nearly one unit if $T/T_r \ll 1$. The over-all result is that lines are spaced at intervals of f_r along the frequency scale, and the ends of the lines follow the envelope curve specified by the factor $(\sin k\pi T/T_r)/(k\pi T/T_r)$ and indicated by the curved line in Fig. 2*B*.

The envelope curve, which has the form of $\sin x/x$, comprises positive and negative loops because of the oscillation of $\sin x$, and the peak amplitude of the loops is less at higher frequencies because of the $1/x$ factor. In the frequency range where the envelope curve is negative, the cosine terms of Eq. 6 have negative coefficients; that is, the cosine waves take on peak negative values instead of peak positive values at $t = 0$. The frequencies separating the positive and negative loops of the envelope curve are those for which $\sin k\pi T/T_r$ is zero, or

$$k\pi \frac{T}{T_r} = \pi, \qquad k\pi \frac{T}{T_r} = 2\pi, \qquad k\pi \frac{T}{T_r} = 3\pi, \qquad \ldots$$

Because $1/T_r$ is the pulse repetition frequency f_r, these expressions may be written

$$kf_r = \frac{1}{T}, \qquad kf_r = \frac{2}{T}, \qquad kf_r = \frac{3}{T}, \qquad \ldots$$

and therefore, kf_r being the frequency of the kth harmonic, $1/T$, $2/T$, $3/T$, . . . are the frequencies at which the envelope curve crosses the horizontal axis.

The relationships between the parameters of the pulse of Fig. 2*A* and the frequency spectrum of Fig. 2*B* may be summarized as follows:

1. The pulse height E does not appear in Fig. 2*B* because E determines only a scale factor for the amplitude of the harmonics, and Fig. 2*B* is a plot of relative amplitudes.

2. The repetition frequency f_r determines the spacing of the lines representing the Fourier components but has no effect on the envelope curve drawn through the ends of the lines.

3. The pulse shape determines the shape of the envelope curve; for a rectangular pulse the envelope always has the form of $\sin x/x$.

4. The pulse duration T determines the spacing of the zeros of the envelope curve.

5. The number of harmonic components within each loop of the envelope curve is T_r/T. (Because the spacing of the zeros is $1/T$, and that of the lines is $f_r = 1/T_r$.)

The validity of the Fourier series for a rectangular pulse can be verified by point-by-point addition of the harmonic components. Many cosine waves must be summed, however, especially when the pulse duration is very short relative to the repetition period. If the ratio T/T_r is small the amplitude of each harmonic component is a tiny fraction of the pulse height, because the factor T/T_r appears in front of the parenthesis in Eq. 6. Any one component, therefore, contributes a negligible amount to the sum of the series, and many components must be added before the sum bears a resemblance to the pulse.

Consider, for example, a 1-μs video pulse having a repetition period of 1,000 μs, as indicated in Fig. 3. According to Eq. 6, the average value

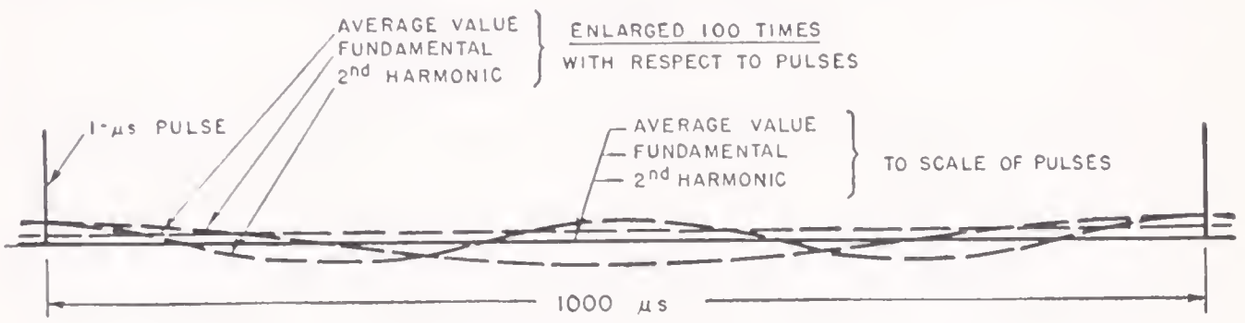


FIG. 3. Three components of 1,000-cps, 1-μs video pulse.

is 1/1,000 of the pulse height, and the amplitude of each of the first few harmonic components is essentially $\frac{1}{500}$ of the pulse height. These components are too small to be visible in Fig. 3 if plotted to scale. When enlarged 100 *times*, the average value, fundamental, and second harmonic have the appearances indicated by dashed lines in the figure. It may be observed that the components, when added, tend to cancel in the interval between pulses and to add directly at the instants of the pulses. When 500 to 1,000 components are summed, nearly complete cancellation results in the between-pulse interval, and the sum at the pulse time approximates the height of the pulse.

To see the effect of adding sine waves in an example requiring fewer additions, consider the pulse in Fig. 4 of 1-μs duration and 8-μs repetition period. In Fig. 4*A*, the sum of the d-c component and the first four alternating components is plotted. The sum includes half the components under the first loop of the spectrum envelope ($T_r/T = 8$) and is therefore equivalent to the sum of 500 components of the 1-μs, 1,000-cps pulse. When four more harmonic components are added to the first sum (Fig. 4*B*), the height of the resulting pulse is increased and the cancellation in the between-pulse interval is improved. The next eight harmonics, added in Fig. 4*C*, belong to the second loop of the envelope curve

and are cosine waves of reversed polarity. Thus the value of the sum at $t = 0$ is decreased, and the double-peaked pulse of Fig. 4*C* is produced.

4. Sine-wave Components of Carrier-frequency Pulses. Fourier-series analysis may also be used to determine the sinusoidal components of the carrier-frequency pulse of Fig. 5*A*. Such a pulse might appear in

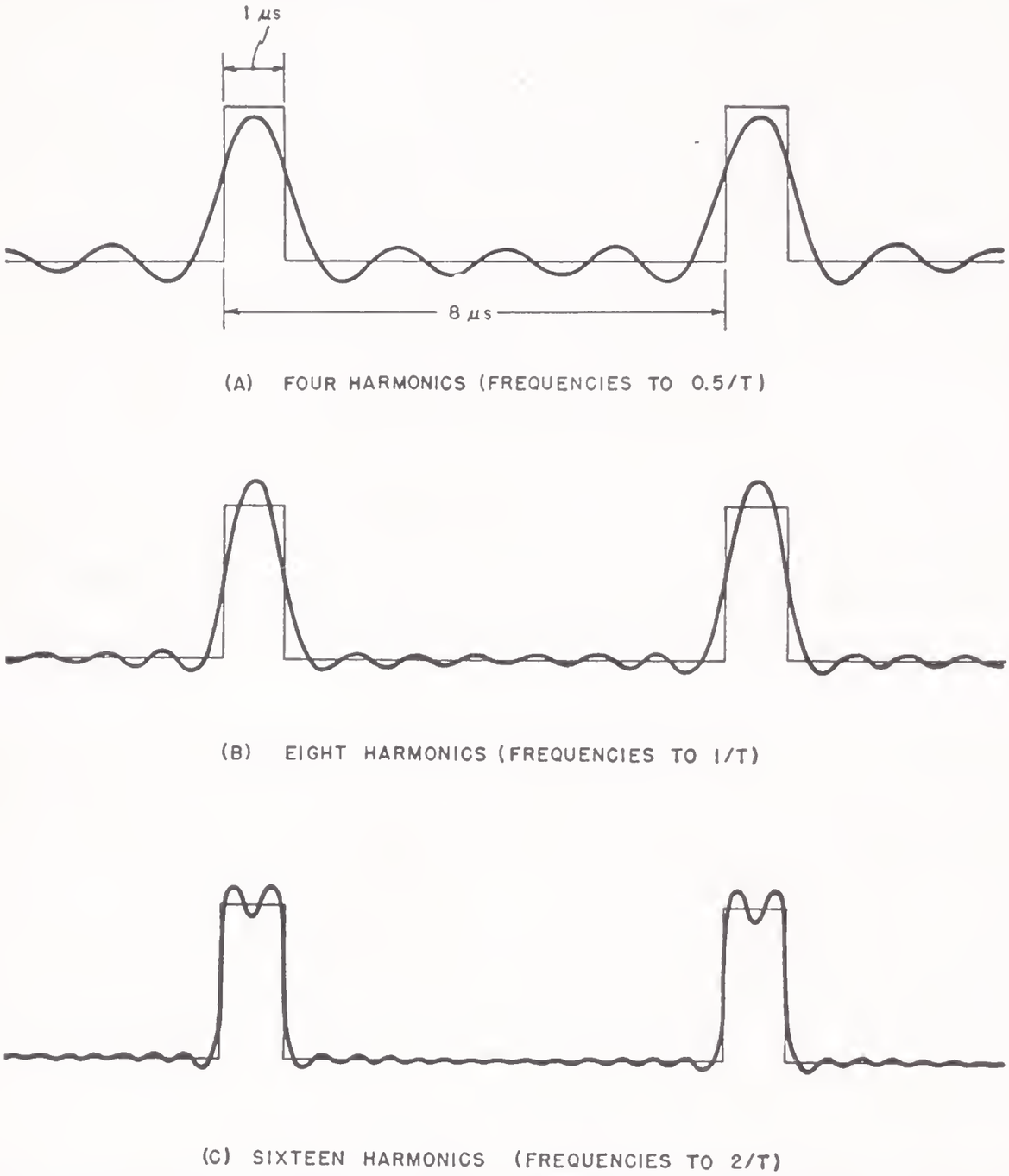
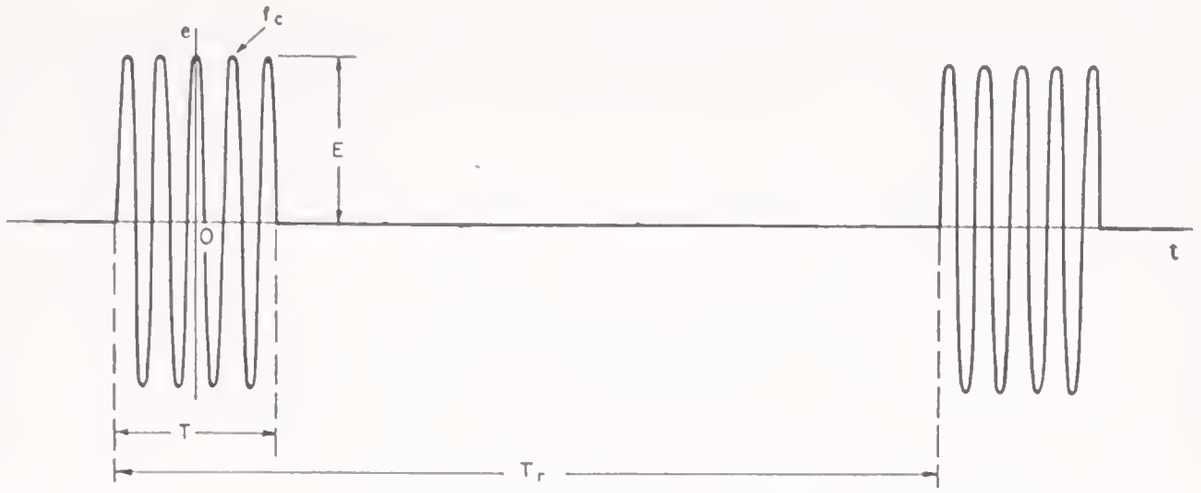
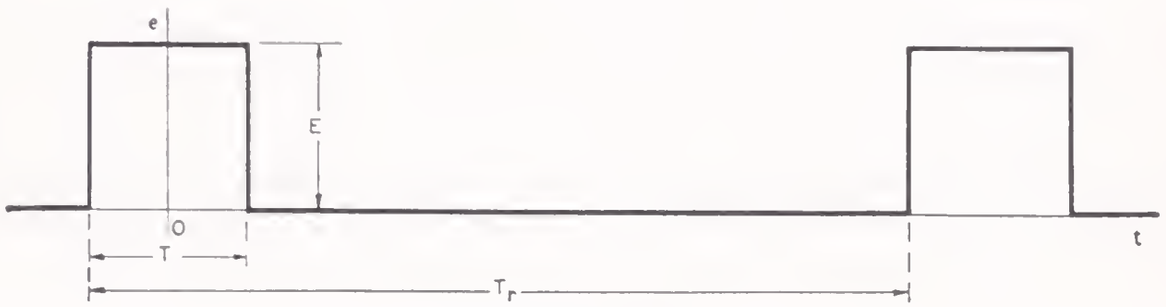


FIG. 4. Comparison of video pulse (light line) with sum of first terms of Fourier series (heavy line).

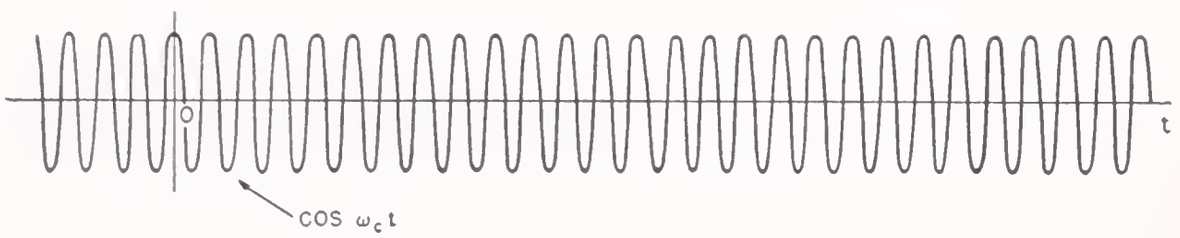
the input circuit of a radar receiver with a carrier frequency f_c of 200 to 30,000 Mcps or in the i-f amplifier with a carrier frequency of 15 to 60 Mcps. For an r-f pulse, the number of carrier-frequency cycles in the pulse would ordinarily be hundreds or thousands of times greater than the number in the figure. If the pulse represents the echo from an object, the ratio T_r/T of repetition period to pulse duration is of the order of 1,000, as for the video pulse in Fig. 3.



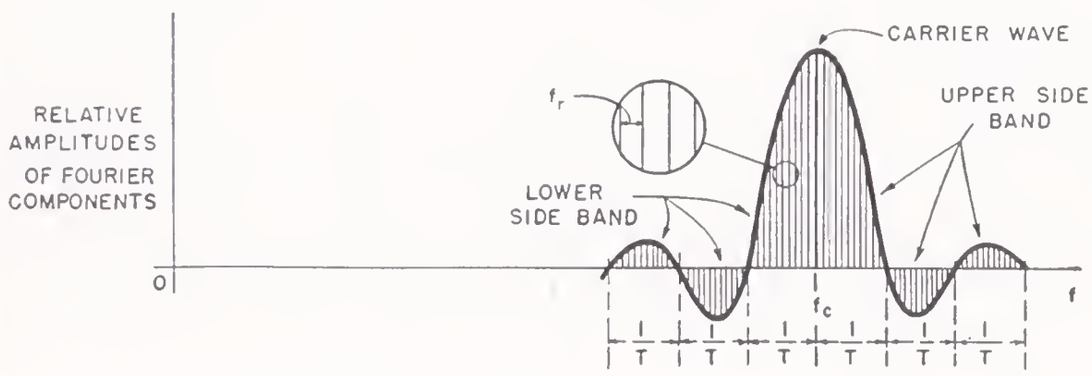
(A) CARRIER-FREQUENCY PULSE



(B) VIDEO PULSE



(C) UNIT-AMPLITUDE COSINE WAVE



(D) FREQUENCY SPECTRUM

FIG. 5. Fourier components of carrier-frequency pulse.

The results obtained in the analysis of a video pulse may be used in the present example if the carrier-frequency pulse is regarded as the product of a video pulse (Fig. 5*B*) and a unit cosine wave (Fig. 5*C*). Multiplying instantaneous amplitudes of the waveforms of Figs. 5*B* and *C* yields a cosine wave of peak value E during the pulse interval and zero in the interval between pulses.

The Fourier series for the video pulse is given by Eq. 6, p. 343. Therefore, an equation for the carrier-frequency pulse is the product of Eq. 6 by a unit cosine wave of carrier frequency f_c and angular frequency $\omega_c = 2\pi f_c$. Thus

$$e = 2E\frac{T}{T_r}\left(\frac{1}{2} + \frac{\sin \pi T/T_r}{\pi T/T_r}\cos \omega_r t + \frac{\sin 2\pi T/T_r}{2\pi T/T_r}\cos 2\omega_r t + \cdots\right)\cos \omega_c t$$

This relation does not show directly the Fourier components of the pulse because the series is the sum of products of two cosine waves instead of the sum of cosine waves. Each product may be separated into two cosine waves through application of the following identity:

$$\cos \theta \cos \phi = \tfrac{1}{2}\cos (\theta + \phi) + \tfrac{1}{2}\cos (\theta - \phi)$$

The resulting Fourier series for e is

$$e = E\left[\frac{T}{T_r}\cos \omega_c t + \frac{\sin \pi T/T_r}{\pi T/T_r}\cos (\omega_c + \omega_r)t + \frac{\sin 2\pi T/T_r}{2\pi T/T_r}\cos (\omega_c + 2\omega_r)t + \cdots + \frac{\sin \pi T/T_r}{\pi T/T_r}\cos (\omega_c - \omega_r)t + \frac{\sin 2\pi T/T_r}{2\pi T/T_r}\cos (\omega_c - 2\omega_r)t + \cdots\right] \quad (7)$$

The frequency spectrum of the carrier-frequency pulse, specified by Eq. 7, is drawn in Fig. 5*D*. The first term of Eq. 7, which corresponds to the average component of the video pulse, is a cosine wave of frequency f_c and is called the *carrier-frequency component.* The remaining terms make up the *side bands.* The upper side band comprises a set of waves spaced at frequency intervals f_r above f_c, and the lower side band comprises a similar set spaced at intervals f_r below f_c. The sin x/x envelope of the video-pulse spectrum appears on each side of the carrier frequency, and the spacing between zero points of the envelope is determined by the pulse duration.

Observe that a carrier-frequency pulse is the sum of Fourier components spread over a band of frequencies, even though the only oscillation observable in the waveform is the carrier oscillation of frequency f_c.

In fact, the spectrum of a carrier-frequency pulse is very closely related to the spectrum of the video envelope of the pulse. To obtain the spectrum of Fig. 5*D* from that of Fig. 2*B*, p. 343, two steps are necessary: (1) On the negative frequency axis in Fig. 2*B*, draw the mirror image of the video spectrum. (2) Move the entire spectrum thus obtained to the right along the frequency axis until all frequencies are shifted by an amount equal to the carrier frequency. This relation between carrier-frequency and video spectra is not restricted to rectangular pulses, but it applies to pulses or waves of any shape, as may be shown by an analysis similar to the derivation of Eq. 7.

5. Sine-wave Response of *R-C* Circuits. An important part of the advantage of Fourier analysis in the study of receivers is the simplification of sine-wave calculations obtained by use of complex numbers. The series *R-C* circuit of Fig. 6*A* is important in the study of video amplifiers. The step-function response of this circuit is explained in Art. 4, Chap. II, and the sine-wave response, obtained with the aid of complex numbers, is presented here.

In Fig. 6*A*, E_i is a complex number standing for e_i, a sinusoidally varying input voltage of frequency f and angular frequency $\omega = 2\pi f$. The magnitude of E_i, denoted by $|E_i|$, represents the rms value[1] of the sine wave, and the angle of E_i represents the phase of the sine wave relative to an arbitrarily chosen reference wave. Suppose that cos ωt is chosen as the reference wave and that the input voltage is

$$e_i = E_{im} \cos (\omega t + \phi)$$

The complex number E_i is then $(E_{im}/\sqrt{2})\epsilon^{j\theta}$. The other voltages and the current of the *R-C* circuit are represented by the complex numbers E_R, E_C, and I, as indicated in Fig. 6*A*.

A convenient representation of the relationships among complex numbers is the complex-number diagram of Fig. 6*B*. For this diagram, the current wave is taken as a reference of phase. Thus complex number I has an angle zero, and the arrow representing I lies along the real axis. The resistor voltage is in phase with the current, and the complex number $E_R = RI$ is represented by a second arrow along the real axis. The capacitor voltage $E_C = I/j\omega C$ lags the current by 90 deg and is represented by an arrow directed downward along the imaginary axis. Addition of E_C and E_R yields the input voltage E_i. This addition is indicated on the diagram. In symbols,

$$E_i = E_R + E_C = RI + \frac{1}{j\omega C} I = \left(R + \frac{1}{j\omega C}\right) I = ZI \qquad (8)$$

[1] Sometimes complex numbers with magnitudes equal to the peak values of sine waves are used. The rms basis is probably more common and is employed throughout this text.

where

$$Z = R + \frac{1}{j\omega C}$$

and Z is the impedance of R and C in series.

If E_R is regarded as the output voltage of the circuit, the ratio E_R/E_i, a complex number, specifies the performance of the circuit. The magni-

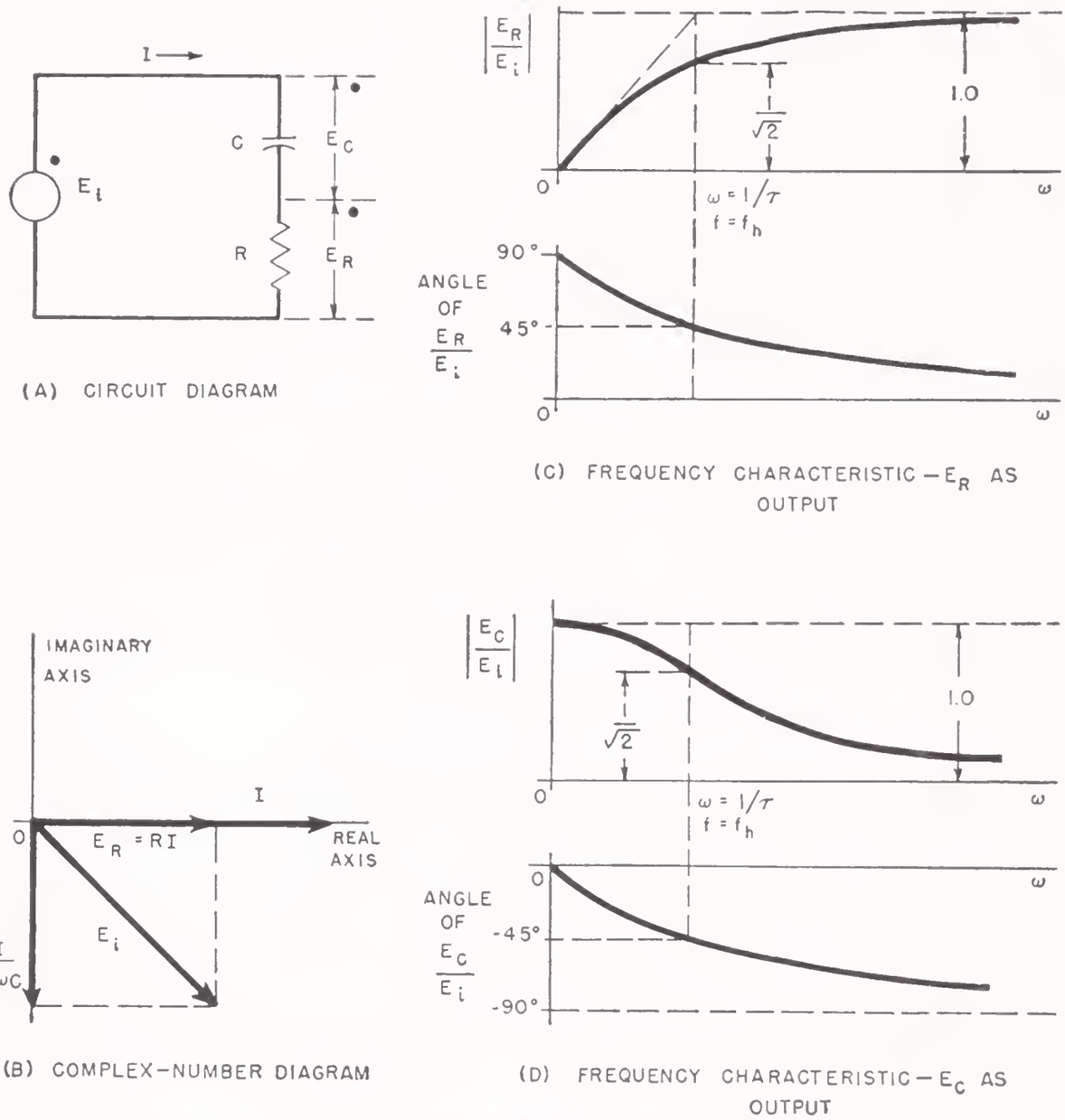


FIG. 6. Sine-wave response of series R-C circuit.

tude of E_R/E_i indicates the effect of the circuit on the rms value of the voltage it transmits, and the angle of E_R/E_i specifies the phase shift produced. Because $E_R = RI$ and, by Eq. 8, $E_i = ZI$,

$$\frac{E_R}{E_i} = \frac{RI}{ZI} = \frac{R}{Z} = \frac{R}{R + 1/j\omega C}$$

$$= \frac{1}{1 + 1/j\omega RC} = \frac{1}{1 + 1/j\omega\tau}$$

where $\tau = RC$ is the *time constant* of the circuit. The magnitude of the complex number E_R/E_i is

$$\left|\frac{E_R}{E_i}\right| = \frac{1}{\sqrt{1 + (1/\omega\tau)^2}} \tag{9}$$

and the angle is

$$\text{Angle of } \frac{E_R}{E_i} = \tan^{-1}\frac{1}{\omega\tau} \tag{10}$$

Equations 9 and 10 indicate that the voltage transmitted depends in magnitude and phase upon the angular frequency ω and the time constant τ. The type of variation with ω is shown by the curves of Fig. 6*C*, obtained by plotting Eqs. 9 and 10. At low frequencies, $1/\omega\tau$ is very large, almost no voltage is transmitted, and the output voltage leads the input by nearly 90 deg. As the frequency increases, the magnitude of the output voltage increases, the leading phase shift decreases, and at frequencies for which $1/\omega\tau$ is very small, the output voltage is nearly equal to the input voltage in both magnitude and phase.

An important parameter of the circuit is the frequency

$$f_h = \frac{1}{2\pi\tau} \tag{11}$$

For $f = f_h$, $\omega\tau$ is equal to unity, the magnitude ratio E_R/E_i is $1/\sqrt{2}$, and the phase shift is 45 deg. Because the power dissipated in R is proportional to the square of $|E_R|$, the power dissipated for $f = f_h$ is one-half of the maximum power dissipated if the magnitude of E_i is held constant as the frequency is varied over the entire range. For this reason, f_h is called the *half-power frequency* of the circuit. The quantity $\omega\tau$ may be replaced in Eqs. 9 and 10 by its equal f/f_h. Thus, if the half-power frequency is known, the response may be calculated immediately from these equations.

In the circuit of Fig. 6*A*, E_C instead of E_R may be considered the output voltage, E_i still being the input. The complex number that specifies the circuit performance is then

$$\frac{E_C}{E_i} = \frac{(1/j\omega C)I}{(R + 1/j\omega C)I} = \frac{1}{1 + j\omega RC} = \frac{1}{1 + j\omega\tau}$$

The magnitude of this complex number is

$$\frac{E_C}{E_i} = \frac{1}{\sqrt{1 + (\omega\tau)^2}} \tag{12}$$

and the angle is

$$\text{Angle of } \frac{E_C}{E_i} = -\tan^{-1}\omega\tau \tag{13}$$

Equations 12 and 13 are used to plot the curves of Fig. 6*D* which show the response of the *R-C* circuit when the capacitor voltage is the output. Note that the output voltage decreases with increase of frequency, is in phase with the input voltage at very low frequencies, and lags in phase at higher frequencies.

The half-power frequency is the same for either choice of output terminals. For the capacitor terminals as output, f_h may be considered to divide the frequency range as follows: Frequencies less than f_h are transmitted with relatively little attenuation and phase shift (see Fig.

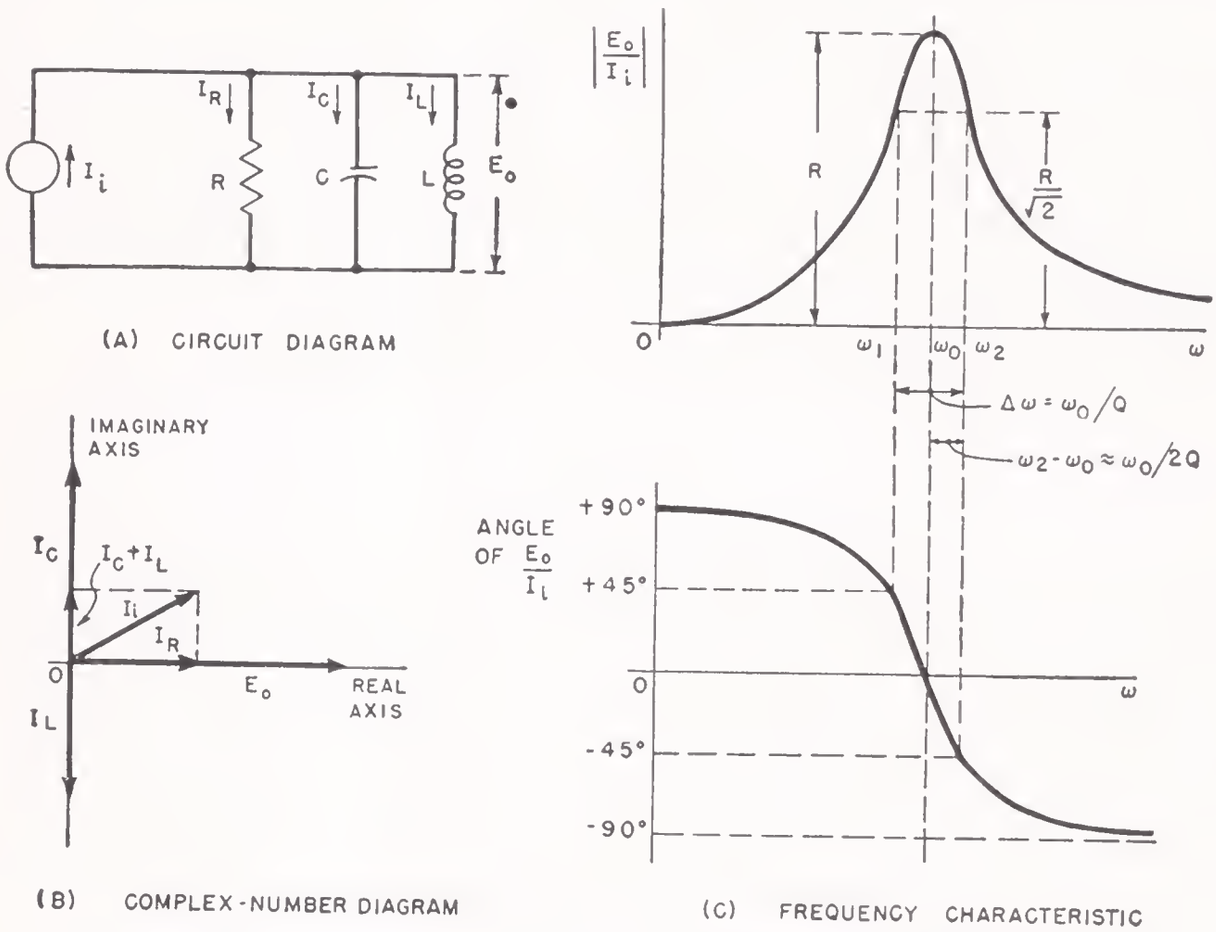


FIG. 7. Sine-wave response of parallel-resonant circuit.

6*D*). Frequencies above f_h are, to a considerable degree, prevented from reaching the output. The circuit may thus be considered to have a *pass band* or frequency range in which signals are transmitted for which the bandwidth is f_h. Of course, the transition between the pass band and the region of rejected signals is not sharp, and for definiteness f_h should be called the *bandwidth to the half-power frequency.*

6. Sine-wave Response of Parallel *R-L-C* Circuit. The circuit of Fig. 7*A* is important in the study of r-f and i-f amplifiers. The transient response of this circuit is discussed in Art. 18 of Chap. II; its steady-state response to sine waves may be determined by complex-number calculations. Let a sinusoidal input current be represented by the complex number I_i, and let the sinusoidal output voltage be E_o. Let currents in

the individual branches of the circuit be represented by I_R, I_C, and I_L, as indicated.

A complex-number diagram for the circuit is drawn in Fig. 7*B*. Voltage E_o is taken as the reference for phase, and E_o and $I_R = E_o/R$ are drawn along the real axis. Capacitance current $I_C = j\omega CE_o$ leads E_o by 90 deg, and inductor current $I_L = E_o/j\omega L$ lags E_o by 90 deg. The sum of the branch currents is the input current. In symbols,

$$I_i = I_R + I_L + I_C = \frac{1}{R} E_o + j\omega CE_o + \frac{1}{j\omega L} E_o$$

$$= \left(\frac{1}{R} + j\omega C + \frac{1}{j\omega L}\right) E_o = YE_o$$

where

$$Y = \frac{1}{R} + j\omega C + \frac{1}{j\omega L}$$

and Y is the admittance of the circuit.

The response of the *R-L-C* circuit is indicated by the complex-number output-input ratio E_o/I_i. This ratio is the reciprocal of Y, or the impedance of the parallel circuit

$$\frac{E_o}{I_i} = \frac{1}{1/R + j\omega C + 1/j\omega L} = R \frac{1}{1 + jR\sqrt{C/L}(\omega\sqrt{LC} - 1/\omega\sqrt{LC})}$$

If the resonant angular frequency $\omega_0 = 1/\sqrt{LC}$ and the sharpness-of-resonance factor $Q = R/\sqrt{L/C}$, defined in Art. 18 of Chap. II, are used here, the impedance relation becomes

$$\frac{E_o}{I_i} = R \frac{1}{1 + jQ(\omega/\omega_0 - \omega_0/\omega)}$$

Thus the magnitude of the response is given by

$$\left|\frac{E_o}{I_i}\right| = R \frac{1}{\sqrt{1 + [Q(\omega/\omega_0 - \omega_0/\omega)]^2}} \tag{14}$$

and the phase shift is

$$\text{Angle of } \frac{E_o}{I_i} = -\tan^{-1} Q\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \tag{15}$$

Plots of Eqs. 14 and 15 are given in Fig. 7*C*. In contrast to the curves of Figs. 6*C* and *D* (which depend upon only one parameter τ), these curves depend upon three parameters—R, ω_0, and Q. Parameter R is a vertical-scale factor, and parameter ω_0 determines the location along the ω axis of the resonant peak. At the peak, $\omega/\omega_0 = \omega_0/\omega$, $|E_o|$ is R times $|I_i|$, and the phase shift is zero. Parameter Q determines the narrowness

of the resonant peak. If Q is large, even a small deviation of ω from ω_0 causes the quantity $Q(\omega/\omega_0 - \omega_0/\omega)$ to be large, and the magnitude of the output voltage is then small and the phase shift nearly 90 deg.

The circuit of Fig. 7*A* has two half-power frequencies, the *lower half-power frequency* f_1 and the *upper half-power frequency* f_2. At both f_1 and f_2, the magnitude of $Q(\omega/\omega_0 - \omega_0/\omega)$ is unity, so that (according to Eqs. 14 and 15) the output voltage is $1/\sqrt{2}$ times its value at resonance, and the phase shift is 45 deg. The half-power frequencies are so located that ω_0 is the geometric mean of $\omega_1 = 2\pi f_1$ and $\omega_2 = 2\pi f_2$, as may be shown from the half-power conditions

$$\begin{aligned} \frac{\omega_1}{\omega_0} - \frac{\omega_0}{\omega_1} &= -\frac{1}{Q} \\ \frac{\omega_2}{\omega_0} - \frac{\omega_0}{\omega_2} &= \frac{1}{Q} \end{aligned} \tag{16}$$

Adding these equations yields

$$\frac{\omega_1 + \omega_2}{\omega_0} - \frac{\omega_0(\omega_2 + \omega_1)}{\omega_1\omega_2} = 0$$

or

$$\omega_1\omega_2 = {\omega_0}^2 \tag{17}$$

Equation 17 states that ω_0 is the geometric mean of ω_1 and ω_2. Furthermore, Eq. 17, written as $\omega_0/\omega_2 = \omega_1/\omega_0$ and substituted into Eq. 16, yields the expression for the separation between half-power points

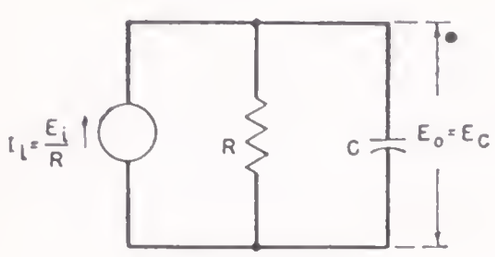
$$\frac{\omega_2 - \omega_1}{\omega_0} = \frac{1}{Q} \tag{18}$$

Thus the angular-frequency bandwidth between half-power points $\Delta\omega = \omega_2 - \omega_1$ is ω_0/Q, the frequency bandwidth Δf is f_0/Q, and the per-unit bandwidth $\Delta f/f_0$ is $1/Q$. If the per-unit bandwidth is small, the geometric mean $\sqrt{\omega_1\omega_2}$ is very nearly equal to the arithmetic mean $\frac{1}{2}(\omega_1 + \omega_2)$, and ω_0 may be considered to be midway between ω_1 and ω_2.

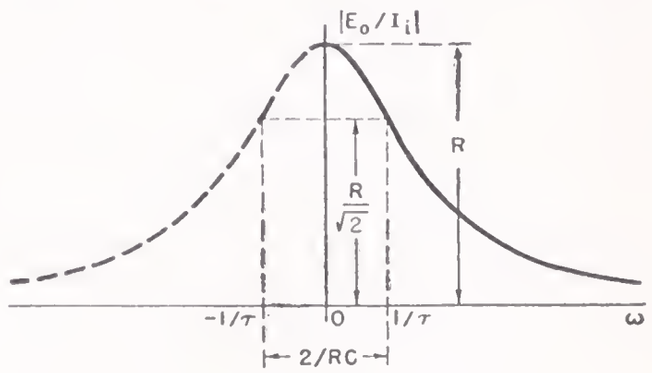
A comparison of the circuit of Fig. 7*A* with the one in Fig. 6*A* (E_C being considered the output voltage) is instructive. These circuits are redrawn in Figs. 8*A* and *B*—the circuit in Fig. 8*A* is equivalent to the one in Fig. 6*A* because Thévenin's theorem may be used to change the current source E_i/R in parallel with R into the voltage source E_i in series with R. The magnitude curve of Fig. 6*D* is multiplied by R and reproduced in Fig. 8*C* as the frequency characteristic of the circuit of Fig. 8*A*. The magnitude curve from Fig. 7*C* is redrawn in Fig. 8*D*. In Figs. 8*C* and *D*, the curves are extended across the zero axis of frequency by plotting points from Eqs. 12 and 14 for negative values of ω. Negative frequencies

have no physical significance; plotting the curves in this way is merely an aid to visualizing the relationships between the responses of the two circuits.

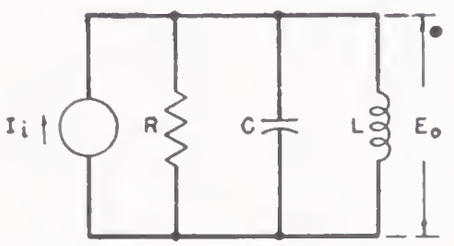
Observe that the circuits of Figs. 8*A* and *B* are alike, except that an inductance L is included in Fig. 8*B*. Addition of the inductance causes the response peak at zero frequency (Fig. 8*C*) to divide into two parts centered at angular frequencies ω_0 and $-\omega_0$ (Fig. 8*D*). The width between half-power points of each of the new peaks is half the corresponding width of the original peak. That is, if R and C have the same values



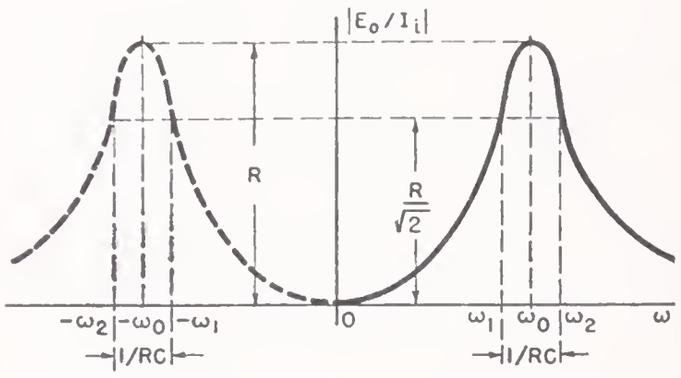
(A) EQUIVALENT CIRCUIT OF FIG. 6A



(C) FREQUENCY CHARACTERISTIC FOR R-C CIRCUIT



(B) CIRCUIT OF FIG. 7A



(D) FREQUENCY CHARACTERISTIC FOR R-L-C CIRCUIT

FIG. 8. Comparison of circuits of Figs. 6*A* and 7*A*.

in the two circuits, the bandwidth of the *R-C* circuit between zero frequency and f_h is the same as the bandwidth of the *R-L-C* circuit between f_1 and f_2. This statement is true because (by Eq. 11, p. 351) $2\pi f_h$ is $1/\tau = 1/RC$ and (by Eq. 18) $2\pi(f_2 - f_1)$ is ω_0/Q, which by the definitions of ω_0 and Q is

$$\frac{1/\sqrt{LC}}{R/\sqrt{L/C}} = \frac{1}{RC}$$

Even if R and C are readjusted so that the bandwidth between half-power points of the response peaks in Figs. 8*C* and *D* are equal, the shapes of the peaks in Fig. 8*D* are not identical with the shape of the peak in Fig. 8*C*. If the per-unit bandwidth of the *R-L-C* circuit is small, how-

ever, the difference in the shapes is negligible. To demonstrate this fact, the quantity $Q(\omega/\omega_0 - \omega_0/\omega)$ in Eqs. 14 and 15 may be written

$$Q\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) = Q\,\frac{\omega^2 - \omega_0^2}{\omega\omega_0} = Q\,\frac{(\omega - \omega_0)(\omega + \omega_0)}{\omega\omega_0}$$

and the approximation $\omega + \omega_0 = 2\omega$ may be made. For a small per-unit bandwidth, this approximation is valid at all frequencies for which the response is appreciable. Thus

$$Q\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \approx 2\,\frac{Q}{\omega_0}\,(\omega - \omega_0) = 2\,\frac{R/\sqrt{L/C}}{1/\sqrt{LC}}\,(\omega - \omega_0) = 2RC(\omega - \omega_0)$$

and substitution of this expression into Eqs. 14 and 15 yields

$$\left|\frac{E_o}{I_i}\right| \approx R\,\frac{1}{\sqrt{1 + [2RC(\omega - \omega_0)]^2}} \tag{19}$$

and

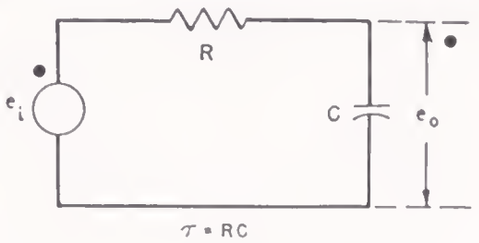
$$\text{Angle of } \frac{E_o}{I_i} \approx -\tan^{-1} 2RC(\omega - \omega_0)$$

Comparison of these equations with Eqs. 12 and 13 ($\omega - \omega_0$ for the resonant circuit being considered analogous to ω for the R-C circuit) shows that the two response characteristics are alike, provided the capacitor in the R-L-C circuit is half as large as the one in the R-C circuit.

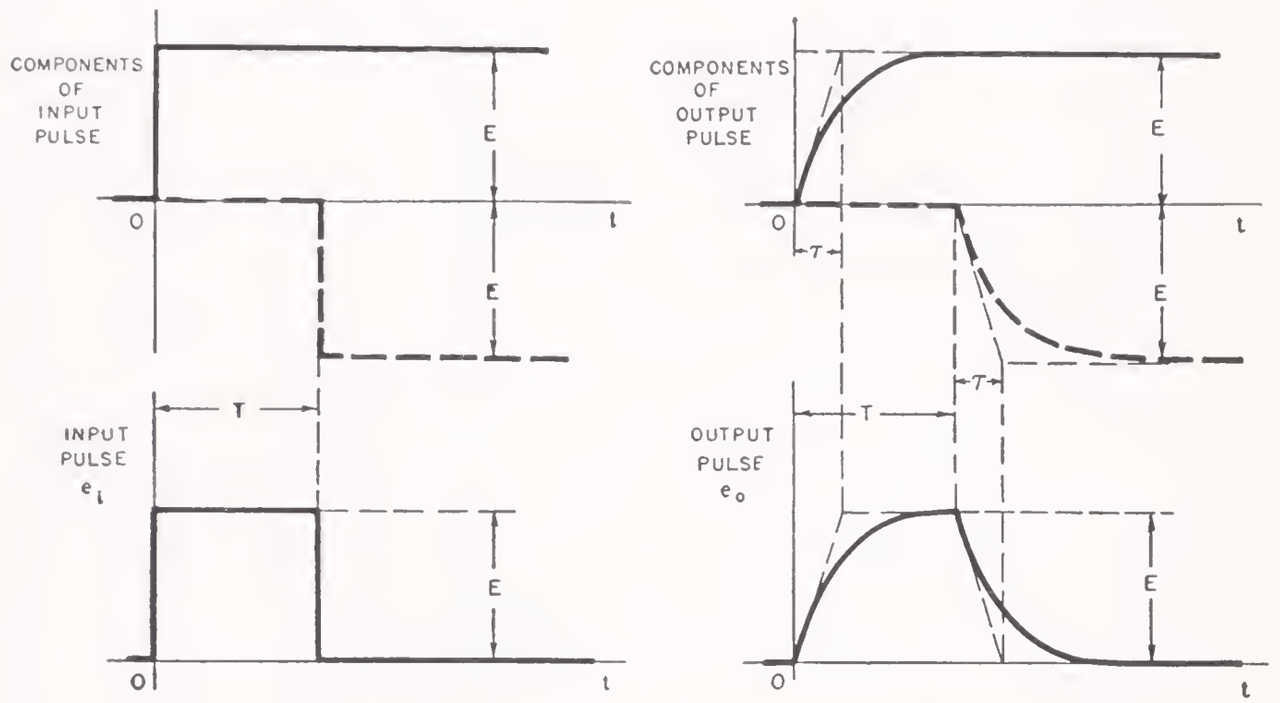
7. Reproduction of Video Pulses. Knowledge of the frequency spectrum of pulses and of the sine-wave response of circuits permits determination of the effect of the circuits on pulse shape. Consider a rectangular video pulse. Its spectrum, drawn in Fig. 2B, p. 343, extends from zero to infinite frequency, and therefore, to reproduce the pulse perfectly, a circuit would be required to respond to all frequencies and to maintain the phase and amplitude relations among all the sine-wave components. Actual circuits are always limited in high-frequency response because of unavoidable shunt capacitances, and usually limitations of low-frequency response also exist. Attenuation of the high-frequency components of pulses affects the reproduction of pulse edges; attenuation of low-frequency components affects the reproduction of long-duration pulses.

The effect of removing the high-frequency components may be seen in one way from the diagrams of Fig. 4, p. 346. In Fig. 4C, only the frequencies above $2/T$ are removed. Since these components form a relatively unimportant portion of the frequency spectrum (see Fig. 2B, p. 343), the pulse reproduction is reasonably good. In Fig. 4B, frequencies above $1/T$ are removed, and the reproduced pulse is less nearly rectangular. Especially, the pulse edges are less steep. When all frequencies above $0.5/T$ are removed (Fig. 4A), a major distortion of the pulse

occurs because of the large number and amplitude of the missing components. The steepness of the pulse edges is greatly reduced, and the pulse is decreased in height and increased in width. The oscillations evident in the reproduced pulses of Fig. 4 can be avoided if high-frequency components are removed by means of a circuit in which the transition between transmitted and rejected frequencies is gradual.



(A) CIRCUIT DIAGRAM



(B) RESOLUTION OF INPUT PULSE INTO STEP-FUNCTION COMPONENTS

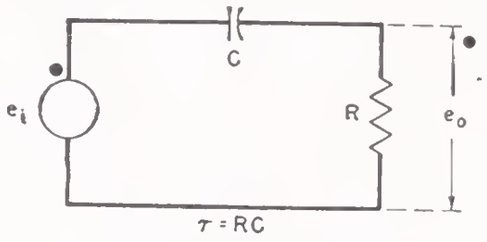
(C) COMBINATION OF EXPONENTIAL CURVES TO FORM OUTPUT PULSE

FIG. 9. Pulse response of R-C circuit (output voltage across capacitor).

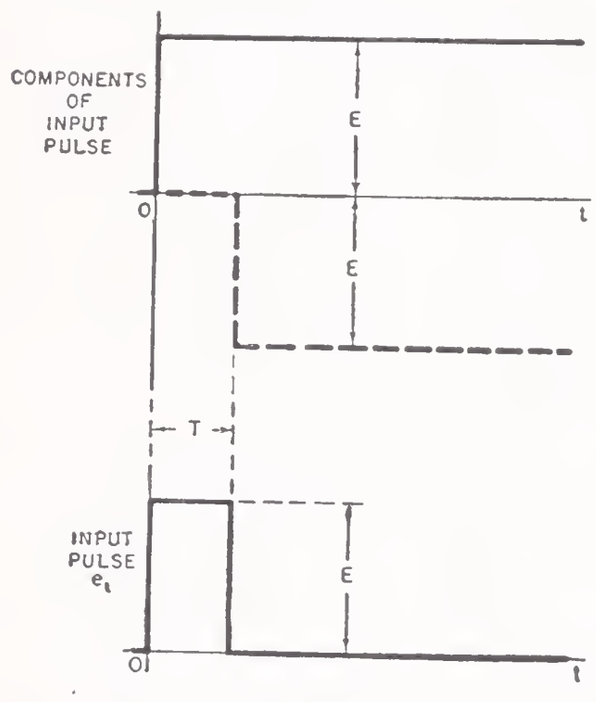
The effect of removing the high-frequency components from video pulses may be seen in another way. The frequency characteristic of the circuit of Fig. 6*A*, p. 350 (output voltage E_c), is given in Fig. 6*D* and indicates that the circuit attenuates high frequencies, the transition between transmitted and rejected frequencies being gradual. The effect on video pulses of such a frequency characteristic can therefore be found by calculating the pulse response of the circuit.

The circuit of Fig. 6*A* is redrawn in Fig. 9*A*, and a rectangular pulse is assumed as the input voltage e_i. In Fig. 9*B*, the pulse is represented as the sum of a positive step function and a delayed negative step function.

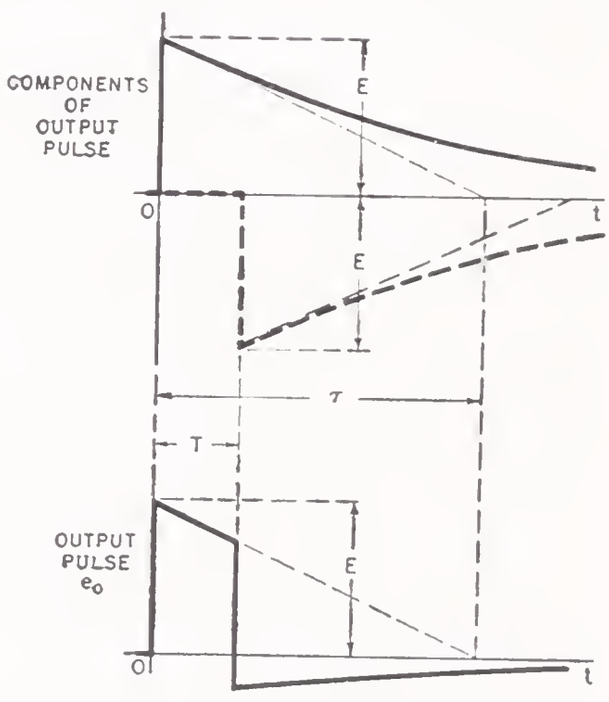
Because the circuit is linear, the pulse response can be calculated as the sum of the responses to the component step functions. These responses are the exponential curves of time constant $\tau = RC$ indicated in Fig. 9*C* (see Art. 4 and Fig. 8, p. 67, in Chap. II), and their sum e_o is the pulse with sloping edges in Fig. 9*C*. The extent to which the pulse edges slope away from the vertical increases with increase of τ or decrease of $f_h = 1/(2\pi\tau)$. Thus the circuit of Fig. 9*A*, by failing to respond to high



(A) CIRCUIT DIAGRAM



(B) RESOLUTION OF INPUT PULSE INTO STEP-FUNCTION COMPONENTS



(C) COMBINATION OF EXPONENTIAL CURVES TO FORM OUTPUT PULSE

FIG. 10. Pulse response of *R-C* circuit (output voltage across resistor).

frequencies, produces a pulse with sloping edges, and the narrower the response band, the more the edges deviate from the vertical. If f_h is so low that τ is comparable to or larger than the pulse duration T, pulse amplitude as well as edge steepness is reduced.

The effect of removing low-frequency components may be seen from a similar calculation. Consider the circuit of Fig. 6*A* (output voltage E_R), redrawn as Fig. 10*A*, for which the frequency characteristic appears in Fig. 6*C*. This circuit attenuates components of low frequency, and the transition between transmitted and rejected frequencies is again gradual.

Calculation of pulse response may be made by use of superposition in the same way as for the preceding circuit. Resolution of the input pulse into step functions is indicated in Fig. 10*B*. The exponential response curves resulting from each step function and the sum of these curves are drawn in Fig. 10*C*. In contrast to the preceding circuit, best reproduction of the pulse requires that the time constant τ be much larger than the pulse duration T. In terms of frequencies, $f_h = 1/(2\pi\tau)$ should be as low as possible, and the value of f_h relative to $1/T$ determines the extent to which the top of the reproduced pulse slopes away from the horizontal. Associated with the sloping pulse top is an overshoot at the end of the pulse. The wider the band of rejected low frequencies, the greater the slope of the pulse top and the more the overshoot.

Video circuits in radar receivers are usually designed so that over a wide band the amplitude response is essentially uniform and the variation in phase shift is negligible. Above some upper limit, however, the response decreases in amplitude and a lagging shift of phase occurs. This upper limit is chosen with regard to the steepness required for the edges of the output pulses. In some applications, steep edges are unnecessary, and the upper limit need be only high enough to prevent appreciable loss of pulse amplitude. Below some lower limit, also, the response of radar video circuits decreases. The lower limit of the pass band is chosen with regard to the duration of the video pulses; inadequate low-frequency response causes a peaking effect in the reproduction of long pulses.

The error is sometimes made of supposing that the fundamental frequency of video pulses must be reproduced and therefore that circuits which respond to the repetition frequency are required. Actually, as long as the pulse duration is short relative to the repetition period, the repetition frequency does not influence the bandwidth requirements for pulse reproduction. This fact is easily seen in terms of the transient calculation. Thus, in Fig. 10, only a single pulse is considered. Effectively, the repetition frequency is zero. Yet the low-frequency response is found adequate if the time constant is long relative to the pulse duration. The point may also be understood in terms of the frequency spectrum of Fig. 2*B*, p. 343. Changing the repetition frequency of the pulses changes the density of the lines in the spectrum but does not change the envelope curve and therefore does not change the relative importance of the various ranges of frequency.

8. Reproduction of Carrier-frequency Pulses. The spectrum of a carrier-frequency pulse, Fig. 5*D*, p. 347, like that of a video pulse, encompasses the entire range of frequencies. Consequently, an ideal circuit that responds equally to all frequencies would be required to reproduce the pulse exactly. Components of frequencies far removed from the carrier frequency f_c, however, are of small amplitude and may be dis-

carded with only a minor effect on pulse shape. Reasonable reproduction of carrier-frequency pulses requires response to a *band* of frequencies centered at f_c. The wider the band, the better is the reproduction of the pulses.

The Fourier-analysis viewpoint indicates that a very narrow-band circuit responding only to the carrier-frequency component cannot reproduce a pulse. Its output would be only the minute sine wave that is the carrier-frequency component in the pulse spectrum. The need for response to frequencies other than the frequency of the oscillations visible in the pulse is sometimes difficult to understand. Consideration of the transient response of tuned circuits to pulses may make the bandwidth requirement more evident.

In Fig. 11*A*, the input current is a carrier-frequency pulse having the waveform shown as i_i in Fig. 11*B*. The frequency of the oscillations within the pulse is f_c, and the resonant frequency of L and C is f_c. The response of the circuit is the voltage e_o, for which waveforms under three different circumstances are drawn in Fig. 11*B*.

First, let L be infinite and C zero; that is, let the entire current i_i be delivered to the resistor R. The voltage e_o is then Ri_i, a carrier-frequency pulse differing from the current pulse only by a scale factor. The input pulse is reproduced perfectly.

Second, let L be a large inductance and C a small capacitance. Calculation of the transient that results when the current pulse is applied is somewhat involved, but the general nature of the response may be understood readily without the calculation. If the duration of the current pulse is sufficiently great, the circuit must eventually reach a steady-state resonant condition in which the sum of the inductance and capacitance currents is zero and i_i flows entirely in R. Thus the final voltage e_o is an oscillation of the same amplitude and phase as that obtained when resistance alone comprises the circuit. Time is required, however, for the final condition to be reached. At the initial instant of the pulse, no energy is stored in either the inductance or the capacitance. During each cycle of the oscillation, energy is delivered to the tuned circuit, and the amplitude of the e_o wave increases as the stored energy increases. After the end of the i_i pulse, the e_o oscillations decay gradually as the stored energy is dissipated in the resistor. (Compare the growth of oscillations in a-c resonance charging of modulator networks, Art. 7 of Chap. III, and the decaying oscillations produced by a ringing circuit, Art. 19, Chap. II.)

Third, let L be a small inductance and C a large capacitance. The larger value of C requires that for a given amplitude of the e_o wave more energy must be stored in the resonant circuit. (The stored energy is equal to $\frac{1}{2}Ce_{o_{max}}^2$ when e_o has its peak value $e_{o_{max}}$.) Because of the greater amount of energy required per unit of oscillation voltage, the e_o

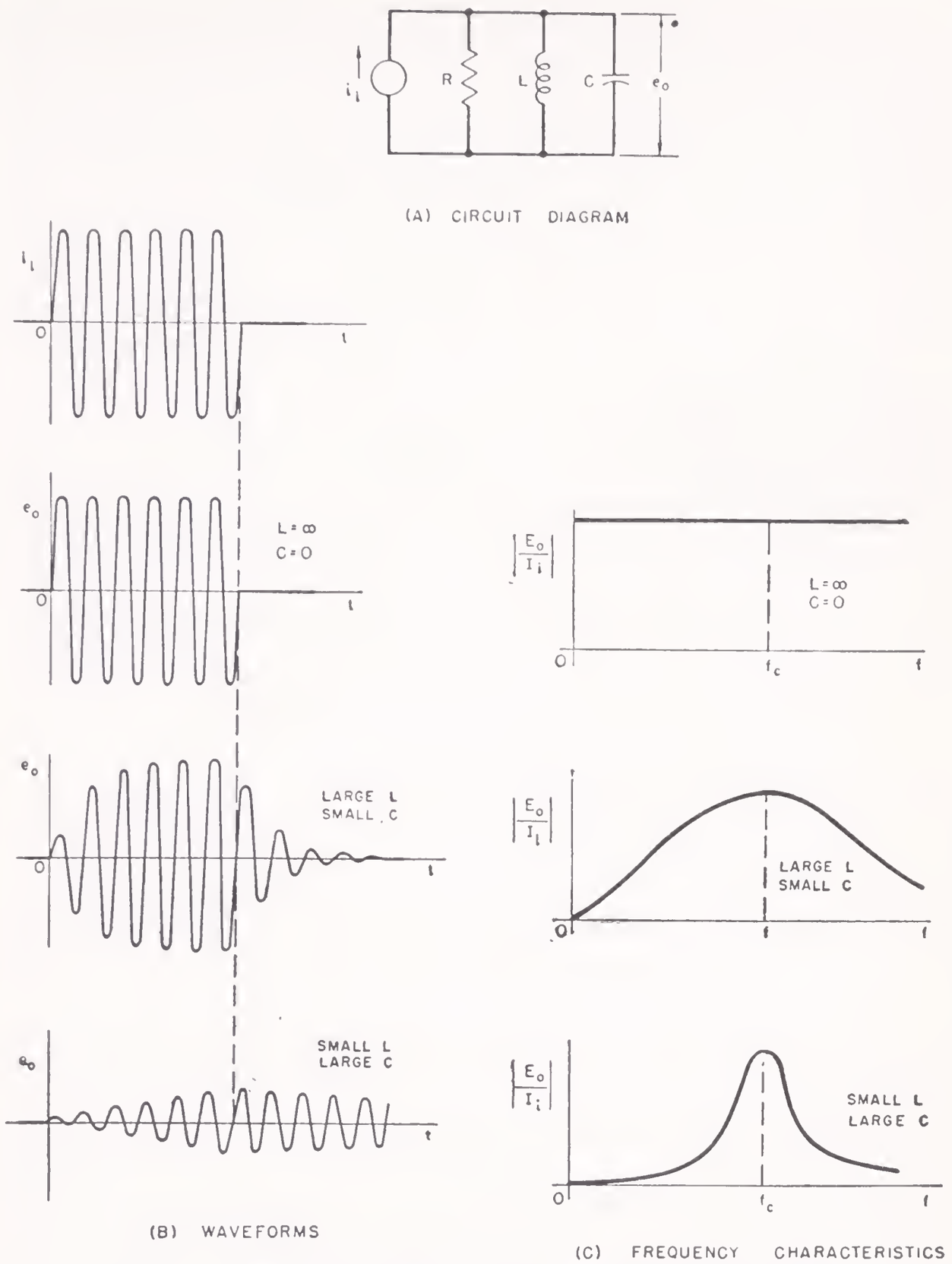
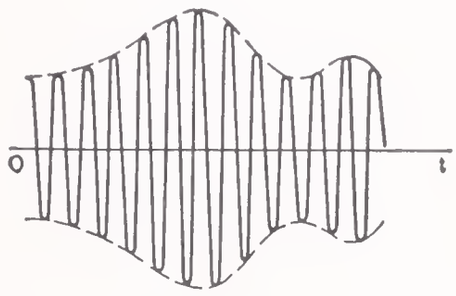


FIG. 11. Response of resonant circuit to carrier-frequency pulse.

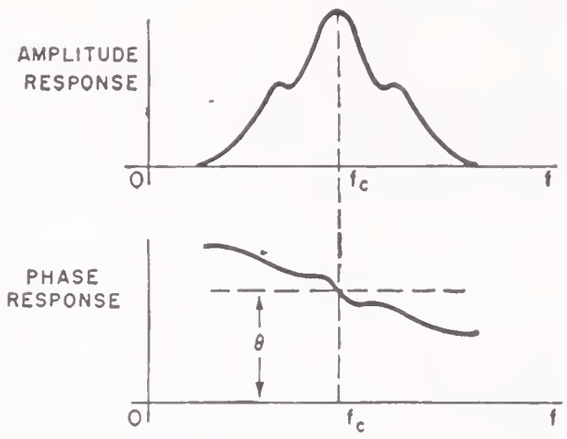
wave grows and decays more slowly, as indicated in the third e_o waveform in the figure.

For the parallel R-L-C circuit, Q is equal to $R/\sqrt{L/C}$. Therefore the three waveforms of e_o in Fig. 11B are in the order of increasing Q. Perfect pulse reproduction requires a circuit of zero Q; nonzero values of Q cause the pulse edges to slope, the distortion growing worse as Q increases. The value of Q determines the per-unit bandwidth, as indicated in Fig. 11C.

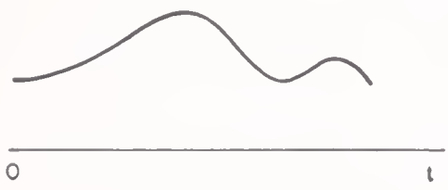
Comparison of the frequency characteristics of this figure with the pulses of Fig. 11*B* shows that removing frequencies remote from f_c has the same effect upon carrier-frequency pulse envelopes as does the removal of high frequencies on video pulses. For a given quality of reproduction, carrier-frequency circuits must provide twice the bandwidth of video circuits in order to accommodate both upper and lower sidebands.



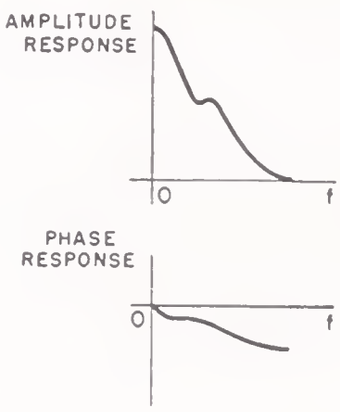
(A) MODULATED CARRIER WAVE



(B) RESPONSE OF CARRIER-FREQUENCY CIRCUIT



(C) VIDEO WAVE



(D) RESPONSE OF VIDEO CIRCUIT

Fig. 12. Input waves and frequency characteristics for related carrier-frequency and video circuits.

The similar effect of removing components remote from f_c and of removing high-frequency components is but one example of a general principle that relates certain carrier-frequency circuits to video or audio circuits.[1] Consider any modulated carrier wave (illustrated by Fig. 12*A*) applied to a circuit having a frequency characteristic of the type illustrated in Fig. 12*B*. The amplitude response is symmetrical about the carrier frequency, and the phase shift is the angle θ plus an angle that varies symmetrically, with sign reversal, about the carrier frequency. Consider also a video wave (Fig. 12*C*) identical with the modulation envelope of the first wave. The video wave is applied to a circuit having

[1] Another example is the relation between a-c and d-c lead networks (Art. 22, Chap. V).

the frequency characteristic of Fig. 12*D*. This figure is obtained from Fig. 12*B* by shifting the origin of f to f_c and by subtracting the angle θ from the phase curve. The general principle states that the modulation envelope of the output carrier wave is identical with the output video wave. The angle θ appears as a phase shift of the carrier-frequency oscillations within the modulation envelope.

Proof of this principle makes use of the relation between the spectra of carrier-frequency waves and of their modulation envelopes (see Art. 4). Let the Fourier series for the input video wave or for the modulation envelope of the input carrier wave be

$$e_{v1} = E_o + E_{m1} \cos (\omega_r t + \phi_1) + \cdots$$

where ω_r is the angular frequency of the fundamental component. If A_n is the factor by which the video circuit multiplies the amplitude of the nth harmonic component of the input wave, and if θ_n is the phase shift of the nth harmonic, the output video wave is

$$e_{v2} = A_o E_o + A_1 E_{m1} \cos (\omega_r t + \phi_1 + \theta_1) + \cdots$$

The input modulated carrier wave is

$$\begin{aligned} e_{m1} &= [E_o + E_{m1} \cos (\omega_r t + \phi_1) + \cdots] \cos \omega_c t \\ &= E_o \cos \omega_c t + \frac{E_{m1}}{2} \cos [\omega_c t + (\omega_r t + \phi_1)] + \cdots \\ &\quad + \frac{E_{m1}}{2} \cos [\omega_c t - (\omega_r t + \phi_1)] + \cdots \end{aligned}$$

The output modulated-carrier wave, because of the phase and amplitude relations specified in terms of Figs. 12*B* and *D*, is

$$\begin{aligned} e_{m2} &= A_o E_o \cos (\omega_c t + \theta) + A_1 \frac{E_{m1}}{2} \cos [(\omega_c t + \theta) + (\omega_r t + \phi_1 + \theta_1)] + \cdots \\ &\quad + A_1 \frac{E_{m1}}{2} \cos [(\omega_c t + \theta) - (\omega_r t + \phi_1 + \theta_1)] + \cdots \end{aligned}$$

or

$$e_{m2} = [A_o E_o + A_1 E_{m1} \cos (\omega_r t + \phi_1 + \theta_1) + \cdots] \cos (\omega_c t + \theta)$$

Comparison of the expressions for e_{v2} and e_{m2} indicates that the modulation envelope of the output carrier wave is identical with the output video wave. Comparison of the expressions for e_{m1} and e_{m2} indicates that the carrier oscillations in the modulated wave suffer a phase shift θ.

This principle is useful in the determination of envelope response for a carrier-frequency circuit if the response of the corresponding video circuit is known. For example, a parallel *R-L-C* circuit of small per-unit bandwidth is shown in Art. 6 to correspond to a video circuit consisting of

resistance and capacitance in parallel. The resistance equals that in the *R-L-C* circuit, and the capacitance is twice the capacitance of the resonant circuit. Thus the principle may be used to show that the edges of the e_o pulses in Fig. 11*B* rise and fall exponentially, the time constant of the exponentials being $2RC$.

9. Linear Equivalent Circuits for Tubes. When input signals to a radar receiver are small so that the amplifier tubes operate as linear devices, the performance of the receiver may be calculated with the aid of equivalent circuits. Each tube is replaced in the receiver circuit by its linear-element equivalent, and complex numbers are employed in the calculation of the sine-wave response of the resulting linear circuit. The nature of the equivalent circuits of tubes depends upon the frequency range considered.

In the range of extremely low frequencies, source-and-resistance equivalent circuits (such as those for the cathode follower, Fig. 6, p. 63 in Chap. II) are adequate. Amplifiers are represented as a current source $g_m e_g$ shunted by a plate resistance r_p or as a voltage source μe_g in series with r_p (g_m is the transconductance and μ the amplification factor of the tube). These circuits are representations of the basic equation for linear operation of an amplifier tube (Eq. 7, p. 64, Chap. II)

$$i_p = g_m e_g + \frac{1}{r_p} e_p$$

where i_p is an increment of the plate current i_b and e_g and e_p are increments of grid voltage e_c and plate voltage e_b, respectively. The two circuits are equivalent and may be used interchangeably for either triode or pentode amplifiers. For the equations and the circuits to apply to pentodes, screen-grid and suppressor-grid voltages must be maintained constant, so that variations of control-grid and plate voltage are the only cause of plate-current variation.

For the range of frequencies present in radar video amplifiers (below about 5 or 10 Mcps), a tube equivalent circuit must include, in addition to the plate resistance and a voltage or current source, three capacitances arranged as in Fig. 13 to account for the interelectrode capacitances of the tube. In the figure, E_g and E_p are complex numbers representing sinusoidal variations of the incremental voltages. At very low frequencies, currents in interelectrode capacitances are negligible, but at frequencies in the megacycle range, these currents often exceed other components of tube current. The circuits of Fig. 13 assume that the total grid voltage e_c is at all times negative so that no electrons reach the grid, and a resistor in parallel with C_{gk} is not required in the equivalent circuit. When the circuits are applied to pentodes, each of the capacitances C_{gk} and C_{pk} represent the sum of several interelectrode capacitances and are usually

called the input capacitance and output capacitance, respectively. Capacitance C_{gp} in pentodes is so small that it may often be neglected, especially at the lower frequencies.

Where frequencies higher than a few megacycles per second are present, as in i-f and r-f amplifiers, the equivalent circuits of Fig. 13 are inaccurate because of *transit-time* and *lead-inductance* effects. Transit time refers to the time of flight of electrons through the tube, and lead inductance to the self and mutual inductances of the connections between internal electrodes of the tube and the external terminals. Though extremely short, less than a hundredth of a microsecond in high-frequency tubes, transit time is appreciable relative to the period of high-frequency signals. Lead inductances also are very small, usually less than 0.1 μh, but they

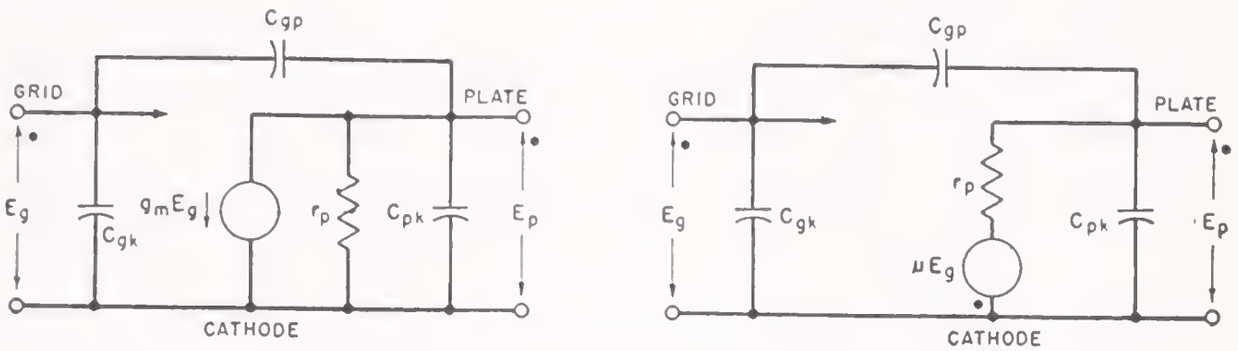


FIG. 13. Linear equivalent circuits for amplifier tubes—interelectrode capacitances included.

have appreciable reactances at frequencies above several megacycles per second.

Another effect of importance at high frequencies is the variation of input capacitance (C_{gk} in Fig. 13) with transconductance. This capacitance is influenced by the stream of electrons passing through the control grid and changes when the transconductance is altered by aging of the tube or by adjustment of gain-control circuits. The capacitance change is essentially independent of frequency, but since it amounts usually to less than a micromicrofarad, its effect is important only at frequencies above the video range.

Lead inductance can be accounted for by the addition of appropriate inductance elements to the circuits of Fig. 13, but transit-time effects cannot be so directly represented. Common practice is to replace the circuit of Fig. 13*A* by the one[1] of Fig. 14*A*. Admittances Y_{gk}, Y_{pk}, Y_{gp}, and y_m are complex-number functions of frequency and are given whatever values are required to represent the tube.

[1] A different equivalent circuit in which less complicated parameter variations are occasioned by transit time is given by F. B. LLEWELLYN and L. C. PETERSON, "Vacuum-Tube Networks," *Proc. I.R.E.*, **32** (March, 1944), 144–166.

The equivalent circuit of Fig. 14*A* may be applied to any linear operating condition of an amplifier tube because it is a circuit representation of the general admittance equations of a two-terminal-pair network. If a voltage E_g and a current I_g are applied at terminal pair 1 (the grid and cathode terminals) and a voltage E_p and a current I_p at terminal pair 2 (the plate and cathode terminals), then

$$\begin{aligned} I_g &= Y_{11}E_g + Y_{12}E_p \\ I_p &= Y_{21}E_g + Y_{22}E_p \end{aligned} \tag{20}$$

where Y_{11} is the admittance measured at 1 with a short circuit at 2, Y_{22} is the similarly defined short-circuit admittance at terminal pair 2, and Y_{12}

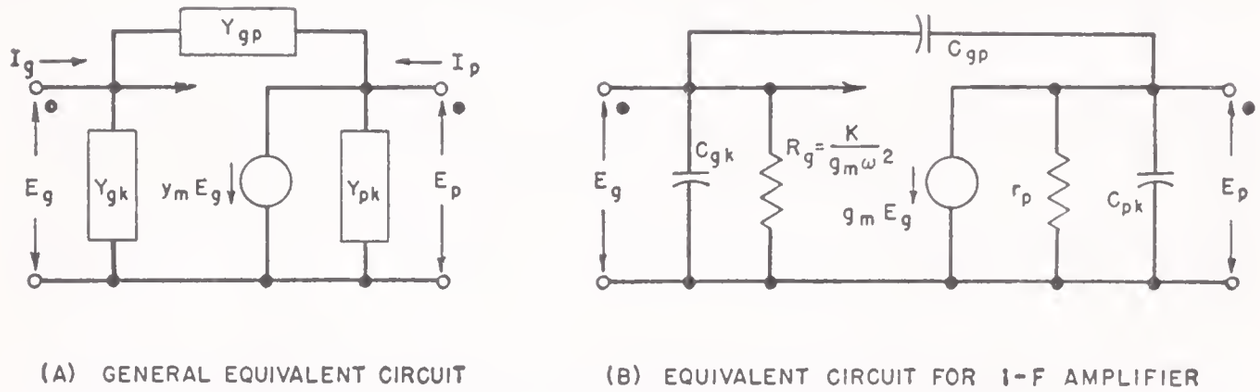


FIG. 14. High-frequency equivalent circuits for tubes.

and Y_{21} are the short-circuit transfer admittance. Applying these definitions to the circuit of Fig. 14*A* yields

$$\begin{aligned} Y_{11} &= Y_{gk} + Y_{gp}, & Y_{12} &= -Y_{gp} \\ Y_{21} &= y_m - Y_{gp}, & Y_{22} &= Y_{pk} + Y_{gp} \end{aligned}$$

For any assigned values of Y_{11}, Y_{12}, Y_{21}, and Y_{22}, corresponding values of the parameters in Fig. 14*A* may be determined from the inverse relations

$$\begin{aligned} Y_{gk} &= Y_{11} + Y_{12}, & Y_{gp} &= -Y_{12} \\ y_m &= Y_{21} - Y_{12}, & Y_{pk} &= Y_{22} + Y_{12} \end{aligned}$$

Over an extended frequency range, each admittance in Fig. 14*A* varies with frequency in a rather complicated manner. Within the limited range of radar intermediate frequencies (about 15 to 60 Mcps), only one difference of the general admittances from the parameters in Fig. 13*A* is of importance: A power loss, called *grid loading*,[1] occurs in the grid circuit even when the grid voltage is negative. Grid loading is the result of both transit time and cathode-lead self-inductance. It is represented in the

[1] W. R. FERRIS, "Input Resistance of Vacuum Tubes as Ultrahigh-frequency Amplifiers," *Proc. I.R.E.*, **24** (January, 1936), 82–107; D. O. NORTH, "Analysis of the Effect of Space Charge on Grid Impedance," *Proc. I.R.E.*, **24** (January, 1936), 108–136.

i-f equivalent circuit of Fig. 14*B* by a conductance $1/R_g$, which is proportional to g_m and to the square of frequency. At 30 Mcps, R_g has a value of the order of 10,000 to 100,000 ohms for useful i-f amplifier tubes.

Additional effects of transit time and lead inductance[1] include a decrease of r_p at high frequency, a dependence of C_{gp} on frequency, and the replacement of g_m by an admittance y_m of magnitude g_m and having a lagging phase angle that increases with frequency. These effects are usually ignored in i-f amplifiers for the following reasons: In the usual pentode amplifiers, r_p is so large relative to the resistances that shunt it

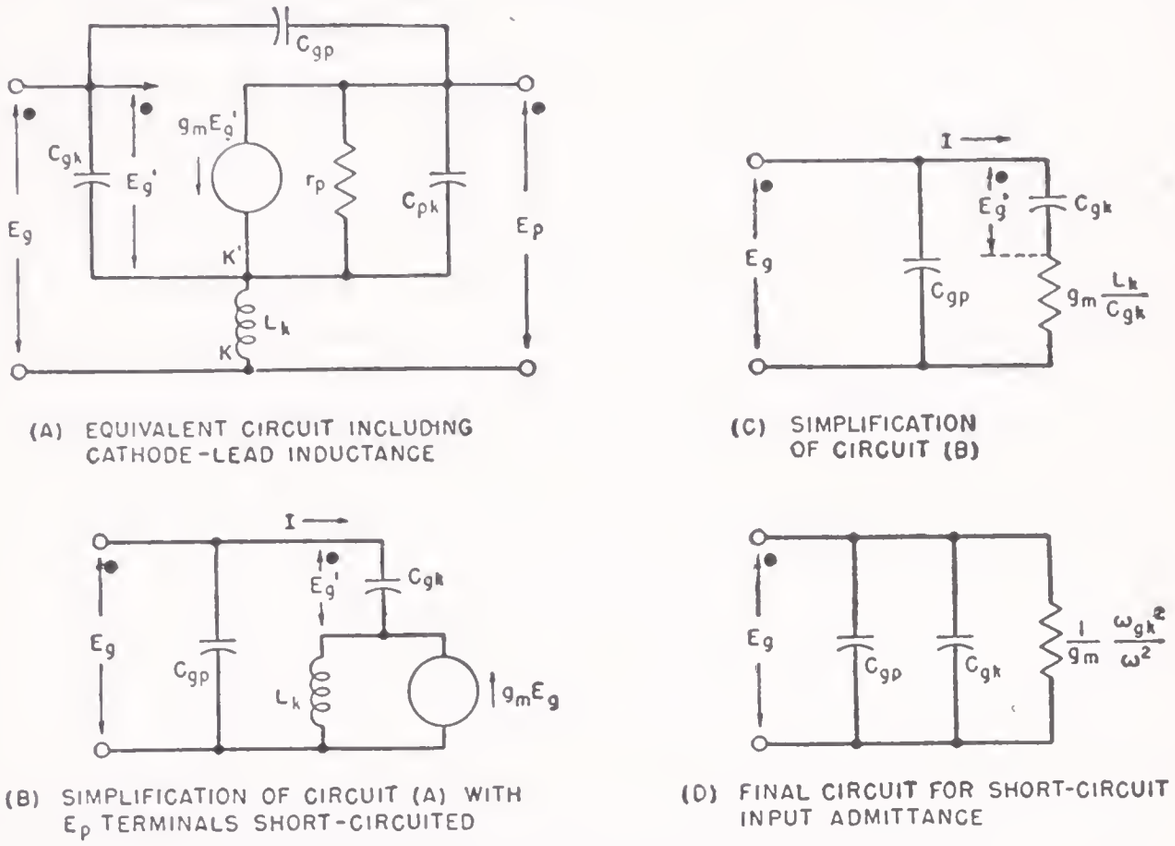


FIG. 15. Grid-loading effect of cathode-lead inductance.

that its exact value is unimportant. Similarly, in pentodes C_{gp} is very small, and the total feedback is determined as much by stray external capacitive and inductive couplings as by C_{gp}. The phase angle of y_m merely causes a minute delay of the output signal, the delay time being of the order of magnitude of the transit time.

10. Grid Loading Caused by Cathode-lead Inductance. To show why inductance of the cathode lead causes an input power loss, an equivalent circuit is drawn in Fig. 15*A*. In this circuit, the only high-frequency effect accounted for is the inductance L_k between the actual cathode (point K') and the external cathode terminal (point K). Voltages E_g and

[1] B. J. THOMPSON, "Review of Ultrahigh-frequency Vacuum-tube Problems," *RCA Rev.*, **3** (October, 1938), 146–155; M. J. O. STRUTT and A. VAN DER ZIEL, "The Causes for the Increase of the Admittance of Modern High-frequency Amplifier Tubes on Short Waves," *Proc. I.R.E.*, **26** (August, 1938), 1011–1032.

E_p are grid and plate voltages measured with respect to the cathode terminal, whereas E_g' is measured with respect to the actual cathode. From Fig. 15*A*, the short-circuit input admittance Y_{11} is to be calculated under the assumption that the reactance of L_k is small but not entirely negligible.

Admittance Y_{11} is the admittance at the terminals of E_g when the terminals of E_p are short-circuited. This short circuit is made in the redrawn circuit of Fig. 15*B*, and r_p and C_{pk} are omitted because they are shunted by the low-reactance element L_k. The current through L_k in Fig. 15*B* is the sum of the currents I and $g_m E_g'$, or

$$\frac{E_g - E_g'}{j\omega L_k} = j\omega C_{gk} E_g' + g_m E_g'$$

and therefore

$$E_g = E_g' \left(1 - \frac{\omega^2}{{\omega_{gk}}^2} + j\omega L_k g_m\right)$$

where ω_{gk} is the resonant angular frequency of C_{gk} and L_k. Because L_k is small, ω_{gk} corresponds to a frequency far above the i-f range, and $\omega^2/{\omega_{gk}}^2$ can be neglected relative to unity. Replacing E_g' by its equal, $I/j\omega C_{gk}$, and dividing by I yields the impedance of the right-hand branch of the circuit in Fig. 14*B*.

$$\frac{E_g}{I} = \frac{1}{j\omega C_{gk}} + g_m \frac{L_k}{C_{gk}}$$

Thus the simpler equivalent circuit of Fig. 15*C* is obtained.

The series *R-C* circuit forming one branch of the circuit in Fig. 15*C* may be changed to a parallel *R-C* circuit if a frequency-dependent resistance is introduced. The change is justified by the calculation of I/E_g, the admittance of the *R-C* branch,

$$\begin{aligned}\frac{I}{E_g} &= \frac{1}{1/j\omega C_{gk} + g_m L_k/C_{gk}} \\ &= \frac{g_m L_k/C_{gk}}{(1/\omega C_{gk})^2 + (g_m L_k/C_{gk})^2} - \frac{1/j\omega C_{gk}}{(1/\omega C_{gk})^2 + (g_m L_k/C_{gk})^2}\end{aligned}$$

The voltage across resistor $g_m L_k/C_{gk}$ in Fig. 15*C* equals the voltage across L_k in Fig. 15*A* and is a small percentage of E_g because L_k is small. Consequently, $g_m L_k/C_{gk}$ is small relative to $1/\omega C_{gk}$, admittance I/E_g becomes

$$\frac{I}{E_g} = g_m \frac{\omega^2}{{\omega_{gk}}^2} + j\omega C_{gk}$$

and the parallel equivalent circuit of Fig. 15*D* is obtained.

The short-circuit input admittance obtained from the video-frequency circuit of Fig. 13 is the parallel combination of C_{gp} and C_{gk}. Figure 15*D* indicates that the effect of cathode-lead inductance is to add in parallel with this capacitance a conductance $g_m\omega^2/\omega_{gk}^2$. This conductance is the parallel combination of the real components of Y_{gk} and Y_{gp} of Fig. 14*A*. Conductance $1/R_g$ in Fig. 14*B*, on the other hand, is the real part of Y_{gk} alone. Calculations based on Fig. 15*A* with $E_p = E_g$ instead of $E_p = 0$ yield the value $(1 + 1/\mu)g_m\omega^2/\omega_{gk}^2$ for the lead-inductance component of $1/R_g$. In most pentode amplifiers, μ is so large that it is not necessary to distinguish between g_m and $(1 + 1/\mu)g_m$.

11. Transit-time Grid Loading. If the total grid voltage of an amplifier tube is at all times negative, none of the electrons passing from cathode to plate strike the grid, and the grid current is entirely the result of changes in the charge stored on the grid. In the determination of the short-circuit input admittance, all electrodes except the grid are assumed connected to the cathode by an a-c short circuit, and the relation between an alternating grid voltage and the rate of change of grid charge it causes is determined.

Assume for the moment that the cathode is unheated so that no electrons occupy the interelectrode space. The grid charge is then the result of capacitances from the grid to all other electrodes. The alternating component of grid charge is $(C_{gk0} + C_{gp0})e_g$, where e_g is the alternating grid voltage and C_{gp0} is the grid-plate capacitance (the subscript 0 designates capacitance values in the absence of space-charge electrons). Capacitance C_{gk0} is the grid-cathode capacitance for a triode, or the sum of grid-cathode, grid-screen-grid, and grid-suppressor-grid capacitances for a pentode.

In actual operation, the cathode is, of course, heated, and electrons are present between electrodes. The grid charge is the charge accounted for by the no-space-charge capacitances plus an additional charge *induced* by the space-charge electrons. Consider, for example, a single electron as it leaves the cathode, travels through the grids, and finally reaches the plate. When close to the cathode, the electron causes a positive charge of the same magnitude as its own negative charge to appear at the adjacent cathode surface. (The electric field surrounding the electron is represented by flux lines emanating from the positive cathode charge and terminating at the electron.) As the electron moves across the tube, the positive charge flows through the external circuit. When the electron is close to a grid wire, the positive charge is mostly on that wire; and when the electron approaches the plate, the positive charge moves to that electrode and is canceled by the electron when it strikes the plate. Let q_e be the alternating component of the sum of all the charges induced on the control grid by all the electrons moving through the tube. The

total alternating grid charge q_g is then the sum of q_e and the alternating charge caused by the no-space-charge capacitance,

$$q_g = q_e + (C_{gk0} + C_{gp0})e_g$$

Because electrons do not strike the grid, the grid current is an alternating current i_g, which is the derivative of the grid charge,

$$i_g = \frac{dq_g}{dt} = \frac{dq_e}{dt} + (C_{gk0} + C_{gp0})\frac{de_g}{dt}$$

This equation for grid current indicates that the input admittance is the parallel combination of the capacitances C_{gk0} and C_{gp0} and another admittance determined by the current dq_e/dt. This additional admittance is the result of the space-charge electrons.

Assume that the alternating voltage e_g is of low frequency so that, relative to the time of a cycle, electrons cross instantly from cathode to plate. The electron current leaving the cathode then increases and decreases in phase with e_g; and the density of electrons at each point in the tube and hence the electron-induced grid charge vary in the same phase. For small changes of electron current i_e, therefore, $q_e = K_e i_e$, where K_e is a constant of proportionality determined by tube geometry and the pattern of space-charge distribution. Because the plate voltage does not vary, i_e equals $g_m e_g$, and thus

$$q_e = K_e i_e = K_e g_m e_g$$

and

$$\frac{dq_e}{dt} = K_e g_m \frac{de_g}{dt}$$

The current dq_e/dt is thus exactly the current that would flow if voltage e_g were applied to a capacitance of value $K_e g_m$. At low frequencies, therefore, space charge has the effect of increasing the input capacitance $C_{gk0} + C_{gp0}$ by the amount $\Delta C = K_e g_m$.

If the frequency of the applied voltage is increased until transit time is of importance, the variations of electron density no longer have the same phase at all points of the tube. An increase of e_g causes an increased number of electrons to leave the cathode. These electrons require time to cross the tube; they pass the grid at a time later than the instant of increase of grid voltage and reach the plate at a still later time. When an alternating grid voltage is applied, the electrons may be visualized as crossing the tube in *bunches*, or regions of increased electron density. Because electrons nearest the grid have most effect in inducing the charge q_e, variations in this charge lag variations in grid voltage by a time proportional to transit time. If the grid voltage is the sine wave

$$e_g = \sqrt{2}\,|E_g| \sin \omega t$$

then the grid charge is

$$q_e = K_e g_m \sqrt{2}\,|E_g| \sin \omega(t - T)$$

where T is the delay time for the grid charge. The grid-current component caused by induced charge is

$$\frac{dq_e}{dt} = \omega K_e g_m \sqrt{2}\,|E_g| \cos \omega(t - T)$$
$$= \omega K_e g_m \sqrt{2}\,|E_g| (\cos \omega t \cos \omega T + \sin \omega t \sin \omega T)$$

If T is a small fraction of the period of e_g, as it is in the frequency range of i-f amplifiers, $\cos \omega T$ is approximately unity and $\sin \omega T$ is approximately ωT, so that the equation for dq_e/dt becomes

$$\frac{dq_e}{dt} = \omega K_e g_m \sqrt{2}\,|E_g| \cos \omega t + \omega^2 T K_e g_m \sqrt{2}\,|E_g| \sin \omega t$$

The induced grid current comprises two components: (1) A current $\omega K_e g_m$ times e_g in magnitude and 90 deg ahead of e_g in phase. This is the current associated with the increment of capacitance $\Delta C = K_e g_m$, which has the same value at high frequencies as at low frequencies. (2) A current $\omega^2 T K_e g_m$ times e_g in magnitude and in phase with e_g. This is the current that would flow in a conductance $\omega^2 T K_e g_m$ proportional to frequency squared and to g_m. This conductance is the grid loading effect of transit time. Figure 16 shows the input admittance separated into no-electron capacitances and the capacitance and conductance effects of space change.

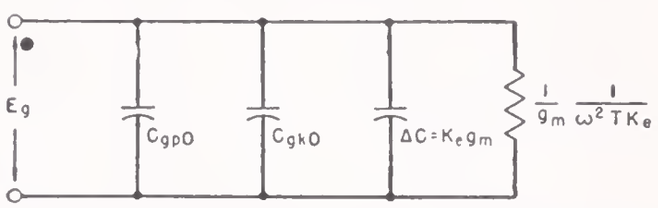


FIG. 16. Short-circuit input admittance of amplifier tube showing effect of transit time at radar intermediate frequencies.

The admittance of Fig. 16 is the short-circuit input admittance, or the sum of the admittances Y_{gk} and Y_{gp} of Fig. 14*A*. The value of Y_{gk} alone may be determined by calculating the grid current for an alternating plate voltage e_p equal to e_g. The current in C_{gp0} is then zero, the electron current leaving the cathode is $(g_m + 1/r_p)e_g$ instead of $g_m e_g$, and the space-charge admittance is multiplied by $1 + 1/\mu$. For high-μ tubes, $1/\mu$ may be neglected; thus elements C_{gk0}, ΔC, and $1/g_m \omega^2 T K_e$ in Fig. 16 may be assigned to Y_{gk}, and C_{gp0} to Y_{gp}. Figure 14*B* therefore represents the combined effects of transit time and lead inductance if $1/R_g$ is the sum of two conductances, each proportional to $g_m \omega^2$, and if C_{gk} includes a small component proportional to g_m.

12. Silicon Crystals. As explained in Art. 1 and indicated in Fig. 1 (p. 339), microwave radar receivers employ *crystal mixers* to change the

r-f input signal to an i-f signal, which is then amplified by vacuum tubes. The silicon crystal in the mixer is a nonlinear rectifying element that may be compared with a diode—its resistance is much less for one direction of current flow than for the other. The rectifying crystals used in mixers should not be confused with the piezoelectric quartz crystals mentioned in Art. 12 of Chap. IV; piezoelectric crystals and rectifying crystals are entirely different devices and perform different functions.

Silicon crystals are mounted in *cartridges* in the manner shown by Fig. 17. The ceramic cartridges of Fig. 17*A* are employed in 10- and 3-cm mixers, and the coaxial cartridge of Fig. 17*B* is used in 1-cm systems. The internal arrangement of the cartridges varies somewhat from one

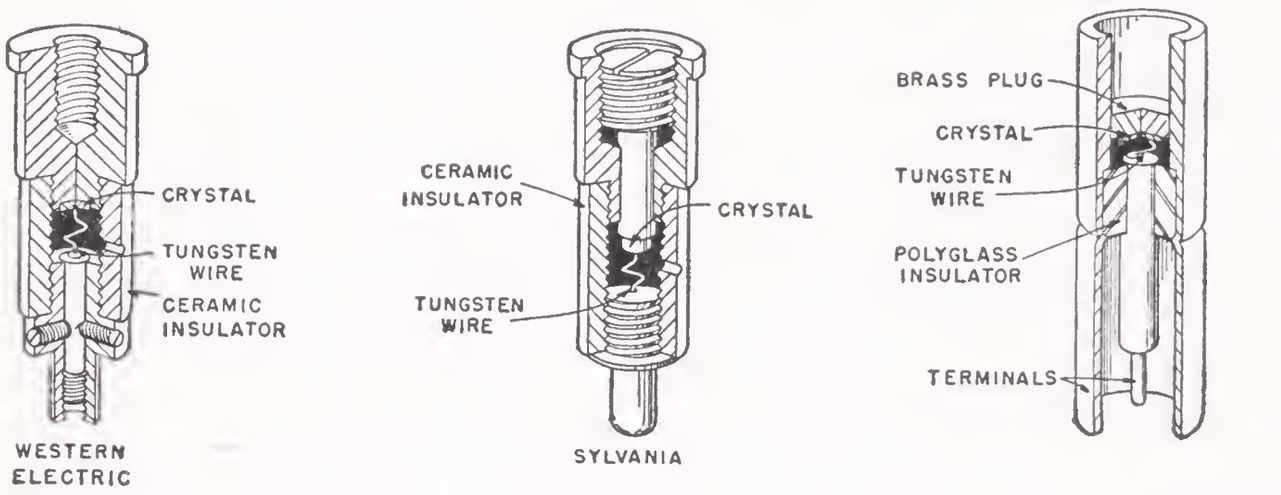


FIG. 17. Cutaway views of crystal cartridges.

manufacturer to another, but the external dimensions have been standardized[1] so that all cartridges of the same type designation are interchangeable. One terminal of a crystal cartridge connects to a fine tungsten wire, and the other terminal supports a crystal of nearly pure silicon. The relative position of the terminals is maintained by an insulating sleeve or bead, and the sharpened point of the tungsten wire is made to press lightly against the surface of the crystal. The impurity content of the silicon and the area of contact at the tip of the tungsten wire (which area depends upon the contact pressure) have an important effect upon the crystal performance and are very carefully controlled in manufacture. A wax filling, which is introduced after the tungsten wire or "cat's whisker" is adjusted, serves to hold the wire in position and to seal the unit against moisture.

The rectifying property of the crystal depends upon the fact that electrons may traverse a boundary between silicon and a metal in one direc-

[1] Standard dimensions may be found, for example, in H. C. TORREY and C. A. WHITMER, *Crystal Rectifiers*, 429–430 (McGraw-Hill Book Company, Inc., New York, 1948).

tion more readily than in the other direction.[1] The boundary between the silicon and its metal support has a large area and negligible resistance to current flow in either direction. The small-area silicon-tungsten contact, on the other hand, has an appreciable resistance. In consequence, the d-c resistance between terminals of the crystal cartridge rises from 100 ohms or less for current flow in one direction to a resistance of the order of 10,000 ohms for current flow in the reverse direction. A typical d-c characteristic is drawn in Fig. 18. The reference directions shown in the insert apply to ceramic cartridges of standard United States construction. For these, current flows most readily if the large terminal of the cartridge is made positive relative to the small terminal.

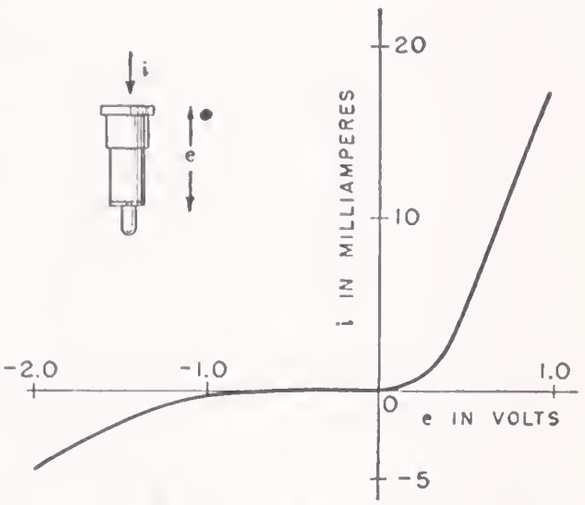


FIG. 18. D-c characteristic of silicon crystal.

An equivalent circuit for a silicon crystal that is an aid to understanding the performance of the crystal in a microwave mixer is drawn in Fig. 19*A*. In this figure, r_b represents the nonlinear resistance of the boundary between silicon and tungsten. Even if no electrons could cross the tungsten-silicon boundary, the crystal would not be an r-f open circuit because of the capacitance between the tungsten point and the silicon. This capacitance, represented by C in Fig. 19*A*, is very small—usually less than 1.0 $\mu\mu$f. The smallness of C and the absence of transit-time effects are reasons why the performance of crystals is superior to that of diodes in microwave mixers. Resistance R in Fig. 19*A* is called the *spreading resistance* of the crystal. It is the resistance of the silicon to current entering from the tungsten point and is appreciable both because silicon is a poor conductor and because the lines of current flow must spread out from the contact point into the body of the silicon. The circuit of Fig. 19*A* represents only the tungsten-silicon contact and the silicon crystal itself. Additional elements must be added if the effects of the inductance of the tungsten wire and

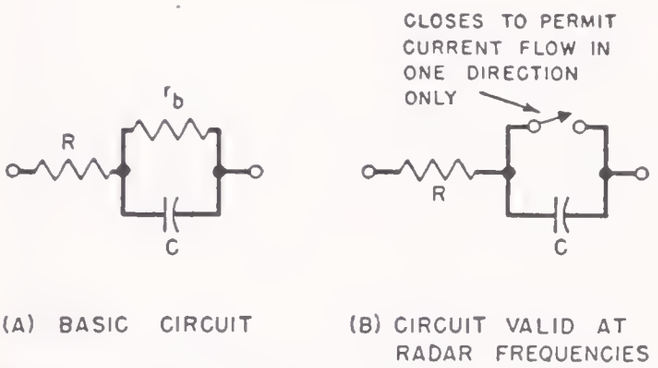


FIG. 19. Equivalent circuits for tungsten-silicon contact.

[1] The theory of conduction in silicon and across the silicon-metal boundary is explained in Torrey and Whitmer, pp. 45–107. A briefer treatment may be found in R. V. POUND, *Microwave Mixers*, 47–54 (McGraw-Hill Book Company, Inc., 1948).

capacitances of the cartridge and the crystal holders are to be accounted for.

A simplified version of Fig. 19*A* is presented in Fig. 19*B*. Because r_b is negligible relative to R for forward voltages greater than a few tenths of a volt, little error is incurred by representing r_b as a closed switch for the forward direction of current. Similarly, at radar frequencies the error produced by assuming r_b to be an open switch when a reverse voltage is applied is negligible—the reactance of C is less than 100 ohms at 3,000 Mcps, and the resistance r_b is at least several thousand ohms. Thus Fig. 19*B* is a valid representation of the crystal for high-frequency operation.

Because at high frequencies the impedance of the crystal to the flow of current in the reverse direction is determined by C and not by r_b, the "front-to-back" ratio (ratio of the d-c resistances for the two directions of current flow) is not a reliable indication of the performance to be expected from the crystal in a radar mixer. As indicated in Fig. 19*B*, the parameters of importance at high frequencies are R and C. The best radar-mixer performance is obtained if R can be made much less than the reactance of C, so that little power is consumed by R, and nearly all the r-f input power is converted to i-f output power. If the contact area between tungsten and silicon has a circular shape, R varies inversely as the radius of the circle and C is roughly proportional to the area of the circle. Thus the smaller the contact area can be made, the smaller is the RC product and the smaller is the ratio of R to the reactance of C. A very small contact area is obtained by use of the sharply pointed tungsten wire to contact the silicon.

The minimum contact area that can be employed in a crystal is determined by the susceptibility of the crystal to damage by excess input power. Power dissipated in the spreading resistance heats and softens the tungsten point and usually causes an increase of the contact area. Crystals may be made very sensitive by use of very small contact areas, but these crystals are easily damaged. Crystals to be used in applications requiring greater ruggedness must have greater contact areas, and sensitivity is sacrificed. Increasing the frequency of operation requires that the contact area be reduced if sensitivity is to be maintained, and thus the higher frequency crystal becomes less rugged.

The power required to damage a crystal is not particularly small, considering the very small size of the crystal parts, yet special care is required in the use of crystals because damage can be done even though the excess power is applied for exceedingly brief periods—periods less than 0.1 μs. The energy that may safely be applied to a crystal by a pulse of less than 1-μs duration is extremely small. Most other circuit elements can be subjected to tremendous overloads if the power is applied only for an instant and can therefore withstand short pulses of considerable energy.

The reason why a crystal can be damaged so quickly is that the parts heated have very small thermal capacity and can change temperature almost instantly.

The possibility that a pulse of brief duration may damage a crystal is a particular disadvantage in a radar system, because a minute fraction of the energy of each transmitted pulse passes through the t-r device (see Fig. 1, p. 339) and reaches the crystal. The t-r device must limit this energy to the minute value that is safe for the crystal. (See Arts. 10 to 14 and 16, Chap. XII, for a discussion of t-r devices and crystal protection.) If the t-r device does not function properly, damage to the crystal usually occurs gradually, and the over-all receiver sensitivity may be reduced greatly before the difficulty is discovered. Frequent checks of the crystal should therefore be made to ensure optimum performance. One procedure for checking a crystal is to substitute a new crystal known to be satisfactory and to note the effect upon receiver performance. An alternate procedure is to measure the "back" resistance (resistance to current flow in the high-resistance direction) when the crystal is first placed in service. The measurement is repeated frequently during use of the crystal, and any appreciable decrease of the back resistance is taken as evidence that the contact area has increased and the crystal performance has deteriorated. If a new crystal is found to deteriorate soon after being placed in service, the t-r device should be checked.

Damage to a crystal when it is not in use is also possible. As protection against damage by strong r-f fields encountered near radar transmitters or other powerful r-f equipment, ceramic crystal cartridges (Fig. 17*A*) should be stored in a shielding container of metal or metal foil. This precaution is unnecessary with the coaxial cartridge, which is self-shielding. When a crystal is inserted into the crystal holder of a mixer, there is a possibility that a static charge on the body may be discharged through the crystal. To avoid damage in this way, the hand should be touched to the crystal holder before the crystal is inserted. Damage may also result from a severe mechanical jar such as may be occasioned by dropping the crystal onto a hard surface. The wax filler supports the tungsten wire, so that a considerable shock is required to disturb the contact. Nevertheless, dropping or other rough treatment of crystals should be avoided.

In Table 1 are listed the type designations of standard mixer crystal cartridges together with certain specifications that indicate the field of application of each type. All types listed, except Type 1N26, make use of the ceramic cartridge sketched in Fig. 17*A*. The 1N26 employs the coaxial cartridge of Fig. 17*B*. The intended wavelength band for each crystal type is indicated in the second column of the table, and the remaining data apply for operation in this band. The figures of the third

and fourth columns together specify the sensitivity of the crystal. Conversion loss and noise-temperature ratio are discussed in Arts. 16 and 28; for the present it is sufficient to know that the conversion-loss figure is a measure of the loss of signal in the crystal and noise-temperature ratio measures the noise generated in the crystal. Low values of both figures indicate a sensitive crystal. The numbers given are specification values that must not be exceeded by new crystals when operated in special test equipment. The next to last column indicates the energy of a video pulse of less than 0.01-μs duration that the crystal is subjected to before being tested for conversion loss and noise-temperature ratio. The tendency for crystals to become less rugged with increase of sensitivity or frequency rating may be observed. The final column is equivalent to a statement of the minimum back resistance of the crystals of each type when new. A back current in excess of the value indicated is a relatively reliable indication that the crystal has deteriorated.

Table 1. Mixer Crystal Cartridges

Type No.	*Wavelength band, cm*	*Conversion loss, db*	*Noise-temperature ratio*	*Test pulse energy, ergs*	*Back current at 1 volt, ma*
1N25	25–30	8.0	2.5		
1N21C	9–11	5.5	1.5	2.0	0.125
1N28	9–11	7.0	2.0	5.0	
1N23A	3–5	8.0	2.7	1.0	0.30
1N23B	3–5	6.5	2.7	0.3	0.175
1N26	1–1.5	8.5	2.5	0.1	0.11* 0.23†

* For cartridges manufactured by Western Electric Co.

† For cartridges manufactured by Sylvania Electric Products Co.

13. Characteristics of Noise. As explained in Art. 4 of Chap. I, the sensitivity of radar receivers is not limited by the amplification attainable but by noise, which obscures weak signals when high amplification is utilized. The nature of noise and the methods of reducing its importance, therefore, have a great influence on receiver design. The input stages, especially, must be constructed with care, because noise entering these stages receives the greatest amplification.

Noise in receivers includes *interference*, *fluctuation noise*, and *avoidable noise*. Interference refers to undesired signals which, along with the desired signal, are received by the radar antenna. Fluctuation noise appears inevitably in all electric circuits, because it is associated with the atomic structure of matter and the charged-particle-in-motion nature of

electric current. Avoidable noise may arise from a great variety of causes, among them, power-supply ripple, vibration of circuit components, faulty contacts, current-carrying composition resistors, gaseous discharges, motor-commutator noise, and stray coupling with adjacent circuits. The effects of all such noise sources may be reduced in any desired degree by proper design of circuits and control of operating conditions. Fluctuation noise and some forms of interference, on the other hand, are fundamental and cannot be reduced below fixed limits.

Fluctuation noise is a minute, irregular, unpredictable variation of voltage and current. Slow and fast variations occur in a random, ever-changing pattern illustrated to some extent by the appearance of noise on a Type A indicator (see Fig. 11A, p. 17, in Chap. I). Both positive and negative excursions of the waveform occur, and over a long period the average value of a noise voltage or current is zero. Because of the randomness of noise, instantaneous values have little significance; yet a measure of noise magnitude is needed. If the instantaneous values of current or voltage are squared, the resulting squared quantity changes erratically but is always positive and has a definite average value over a sufficiently long time. This *average-square value* is used as a measure of the magnitude of the noise. Because average power dissipated in a resistor is proportional to the average-square value of current in the resistor or voltage across the resistor, *average noise power* is also a measure of noise magnitude.

The effect of receiver circuits on noise, like the effect on signals, is conveniently determined by resolving the noise into sinusoidal components and finding the effect on each component separately. Resolution of noise waveforms is complicated by the fact that noise is not periodic, so that Fourier-series analysis cannot be applied directly. Instead, a sample of noise wave in a specified time interval is chosen, and a Fourier series is determined for a wave that is the periodic repetition of this sample. The resulting series depends upon the particular sample analyzed, but it is found that, as the length of the sampling interval is increased, the envelope of the spectrum of amplitudes of sine-wave components approaches a definite limit. This limit is the frequency spectrum of the noise. The phases of the sine-wave components, on the other hand, change indefinitely as the sampling interval increases. Fourier components of noise, therefore, have random, undeterminable phases. Increasing the sampling-interval length causes the individual sine-wave components to be spaced more and more closely along the frequency scale. In the limit, therefore the noise spectrum is continuous; that is, noise contains components of all frequencies.

When two fluctuation-noise voltages are combined in series, the resultant voltage is, of course, the sum *instant by instant* of the component

waves. The sum waveform fluctuates erratically in the same manner as the component waves. If the voltages are obtained from different sources, the *average-square voltage* of the sum wave equals the *sum of the average-square voltages* of the components. An equivalent statement is that the power dissipated in a resistor to which the sum wave is applied is *the sum of the powers dissipated by the component waves applied separately.* When noise currents from different sources are combined in parallel, average-square values of current or average noise powers are added. This law of combination is the reason that average-square value instead of rms value is a convenient measure of noise magnitude.

The rule for combining noise waves differs greatly from the familiar rules for combining direct currents and equal-frequency sine waves. If, for example, direct-voltage sources of magnitudes E_1 and E_2 are placed in series, the total voltage is either $E_1 + E_2$ or $E_1 - E_2$, depending upon relative polarities. The average-square voltage of the sum is therefore either

$$(E_1 + E_2)^2 = E_1^2 + E_2^2 + 2E_1E_2$$

or

$$(E_1 - E_2)^2 = E_1^2 + E_2^2 - 2E_1E_2$$

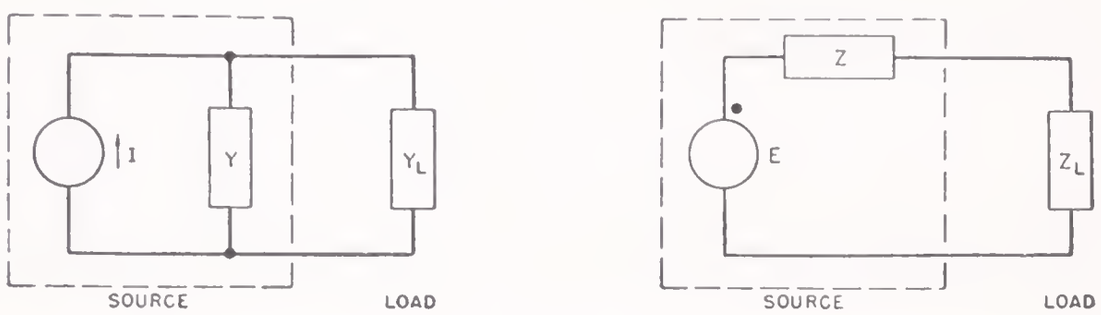
That is, the average-square value of the sum is the sum of the average-square values of the components, plus or minus an *interaction term* $2E_1E_2$. Similarly, if two sine-wave sources of the same frequency are combined, E_1 and E_2 in the above equations may be considered rms values of the component waves, and the interaction term to be added then has any value from $2E_1E_2$ for in-phase waves through zero for waves in quadrature to $-2E_1E_2$ for waves 180 deg out of phase. With noise waves from different sources, the interaction term is always zero—a fact which seems reasonable in view of the random phases of the sinusoidal components of noise.

Waves for which the interaction term is zero are said to be *uncorrelated* or *incoherent.* Noise waves from different sources, a noise wave and any periodic waveform such as a sinusoid or a pulse, and sine waves of different frequencies are examples of uncorrelated waves. A correlation or coherence exists between two waves if the interaction term differs from zero. Sine waves of equal frequency and not in phase quadrature are correlated, and a noise wave is correlated with itself or with a larger scale replica of itself.

Fluctuation-noise voltages are of such minute magnitude that only those which arise in the first stages of the receiver are amplified sufficiently to be important. Input-circuit components in which noise originates include (1) linear, passive circuit elements (resistors, transformers, tuned circuits, transmission lines, and waveguides), (2) antennas,

(3) vacuum tubes, and (4) silicon crystals. The fluctuation noise generated by these elements is discussed in the following articles.

Circuits of the receiver beyond the first stage or two, though not sources of noise of appreciable magnitude, do influence the noise limitation of signal discernibility. Because noise and signal are very different in character, amplifying, detecting, and indicating systems may to some degree operate upon the two types of waves differently. In linear i-f and video amplifiers, distinction between signal and noise can result only from bandwidth restrictions. The effect of bandwidth is discussed in Art. 18. In detectors, indicators, or other nonlinear devices that may be attached to a receiver, effects of a very complicated nature can occur. A few possibilities are mentioned in Arts. 26 and 31.



(A) CURRENT-SOURCE EQUIVALENT (B) VOLTAGE-SOURCE EQUIVALENT

FIG. 20. Equivalent circuit for linear sine-wave source connected to linear load impedance.

14. Noise in Linear Passive Circuits. A source of noise in a *linear* system, like a signal source, may be represented as a current source shunted by an admittance (Fig. 20*A*) or as an equivalent voltage-source circuit (Fig. 20*B*). In passive systems, there are no sources of power other than noise, and I and E may be considered complex numbers representing a single sine-wave component of noise. Correspondingly, complex numbers Y and Z are source admittance and impedance at the frequency of the component considered, and Y_L and Z_L are complex numbers representing the admittance and impedance of a load that may be connected to the noise source.

As a preliminary to the discussion of noise magnitudes, the adjustment of load impedance required for a source to deliver maximum power and the amount of this maximum power are of interest. Consider Y and Y_L in Fig. 20*A* to be separated into conductance and susceptance components. Susceptance current contributes nothing to the power absorbed by the load but does cause a reduction of load voltage. For maximum load power, therefore, the susceptance current should be zero; that is, the load susceptance should be made equal to the negative of the source susceptance. With this condition met, I divides between the source conductance G and the load conductance G_L, and the load power P is

$$P = |E_{\text{load}}|\,|I_{\text{load}}| = \frac{|I|}{G_L + G}\frac{|I|G_L}{G_L + G}$$

or

$$P = \frac{|I|^2}{G}\,\frac{1}{(\sqrt{G_L/G} + \sqrt{G/G_L})^2}$$

where vertical bars denote the magnitudes of the complex numbers they enclose. The first factor in the final expression for P is independent of G_L, and the second factor has a maximum value ¼ when G_L/G is equal to 1. Thus maximum output power is obtained when the load is adjusted so that Y_L is the complex conjugate of Y. This adjustment is often called a *conjugate match*, or the load is said to be *tuned* (for zero total susceptance) and *matched* (for $G_L = G$). The power delivered when the load is tuned and matched,

$$P_a = \frac{|I|^2}{4G} \tag{21}$$

is called the *available power* of the source.

In similar fashion, the available power of the source in Fig. 20*B* may be shown to be

$$P_a = \frac{|E|^2}{4R} \tag{22}$$

where R is the resistance component of Z, and this power is obtained when Z_L is the conjugate of Z. If Figs. 20*A* and *B* represent the same source, G is the real part of $Y = 1/Z$, which is $R/|Z|^2$ or $R|Y|^2$. Also, $I = YE$, and Eq. 21 becomes

$$P_a = \frac{|I|^2}{4G} = \frac{|Y|^2|E|^2}{4R|Y|^2} = \frac{|E|^2}{4R}$$

in agreement with Eq. 22.

Electrons in every conductor have a kinetic energy determined by the temperature of the conductor and because of this energy engage in a chaotic, random motion. The electron motion causes the charge at each point of the conductor to fluctuate erratically, but only by a minute amount. Noise currents and voltages in linear, passive elements result from this fluctuation of charge and are called *thermal noise* or *thermal-agitation noise*. The magnitude and frequency spectrum of thermal noise can be calculated from a very simple relationship that has been derived theoretically[1] and amply verified experimentally.[2] One form of

[1] H. NYQUIST, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, **32** (July, 1928), 110–113.

[2] J. B. JOHNSON, "Thermal Agitation of Electricity in Conductors," *Phys. Rev.*, **32** (July, 1928), 97–109.

this relation, an expression for the available noise power at any terminal pair of a linear, passive network having all its components at the same temperature, is

$$\Delta P_a = kT\Delta f \tag{23}$$

where

Δf is the width of a band of frequencies within which sinusoidal components of noise are considered.

ΔP_a is the available noise power within this frequency band.

T is the temperature of the network in degrees Kelvin (degrees centigrade plus 273).

k is the Boltzmann constant, a basic physical constant relating temperature to energy. Its value is 1.38×10^{-23} joule per degree Kelvin. The quantity kT is an energy, and multiplication of kT in joules by Δf in cycles per second yields ΔP_a in joules per second or watts.

As an illustration of the application of Eq. 23, the open-circuit noise voltage of a resistor may be computed. Let the noise source be a resistor R, so that Z in Fig. 20B becomes R. In that figure, E is the complex rms value of a single sine-wave component of open-circuit voltage for which the average-square value is $|E|^2$. The symbol $\Delta|E|^2$ may be used to denote the sum of the average-square values of all sinusoidal components in the frequency band of width Δf. From Eqs. 22 and 23, therefore,

$$\Delta P_a = \frac{\Delta|E|^2}{4R} = kT\Delta f$$

and

$$\Delta|E|^2 = 4kTR\Delta f \tag{24}$$

Because $4kTR$ is constant, Eq. 24 indicates that the average-square value of the open-circuit resistor noise voltage in a band of fixed width is independent of the center frequency of the band; that is, the frequency spectrum of resistor noise is a horizontal line. Noise of this character is often called white noise. If, for example, R is a 1,000-ohm resistor at a temperature of 17 C, between input terminals of an amplifier having a 2-Mcps pass band, then the input noise voltage that is amplified has an average-square value

$$4 \times (1.38 \times 10^{-23}) \times (273 + 17) \times 1{,}000 \times (2 \times 10^6) = 32.0 \times 10^{-12} \text{ volts squared}$$

or an rms value of 5.65 μv.

A relation similar to Eq. 24 yields the average-square short-circuit noise current obtained from a resistor of conductance G.

$$\Delta I^2 = 4kTG\Delta f \tag{25}$$

Equation 25 is obtained from Fig. 20*A* and Eqs. 21 and 23 in a manner completely analogous to the derivation of Eq. 24.

A somewhat more general example is the calculation of the open-circuit voltage of any uniform-temperature linear network. Equations 22 and 23 apply, as in the resistor example, but R is now the resistance component of the network impedance and may be a function of frequency. With R thus defined, therefore, Eq. 24 applies to any impedance. Similarly, Eq. 25 applies in general if G is the conductance component of the network admittance. If Y in Fig. 20*A* is a conductance, capacitance, and inductance in parallel, the current source has the uniform spectrum associated with constant conductance. The open-circuit noise voltage results from this source current flowing through Y and has a spectrum with the shape of a resonance curve—the average-square voltage is proportional to the square of the impedance of the tuned circuit.

If the components of a network are not all at the same temperature, Eq. 23 may be applied to each of the uniform-temperature parts in turn. The total average-square value of the noise delivered by the network is then the sum of the average-square values of the contributions from the several parts.

Because thermal noise is the result of kinetic energies of electrons determined by temperature, principles of thermodynamics may be expected to apply to noise. Consider resistances R_1 and R_2 at temperatures T_1 and T_2, respectively, with their terminals connected. The noise power delivered by R_1 to R_2 is $mkT_1\Delta f$, where m is a mismatch factor relating the power delivered to the power available from R_1. Calculations similar to those used in the derivation of Eq. 21 show that m is $4/(\sqrt{R_2/R_1} + \sqrt{R_1/R_2})^2$. Because the mismatch factor is unaltered when R_1 and R_2 are interchanged, the noise power that R_2 delivers to R_1 is $mkT_2\Delta f$, and the net transfer of power in the direction R_1 to R_2 is

$$P_{12} = mk(T_1 - T_2)\Delta f$$

The equation for P_{12} is in accord with the second law of thermodynamics. It indicates that noise currents transfer energy from a resistor at high temperature to one at low temperature. In a fictitious system perfectly insulated from its surroundings, transfer of noise power would eventually bring the resistors to thermal equilibrium—a condition of equal temperature and no net power transfer. Such considerations have been used in the theoretical derivation of Eq. 23.

The net power transferred cannot be used in the calculation of noise currents or voltages because the total average-square voltage is the *sum* of the average-square contributions from the two resistor-sources. For example, if resistors R_1 and R_2 are both at temperature T, the net power transferred is zero, but a nonzero noise voltage exists. The power

$mkT\Delta f$ transferred to R_2 produces an average-square voltage $mkTR_2\Delta f$ across R_2—and also across R_1, because the terminals are connected. The same noise power transferred to R_1 produces an average-square voltage $mkTR_1\Delta f$ at the same terminals. The total average-square voltage at the terminals is thus

$$\Delta|E|^2 = mkT(R_1 + R_2)\Delta f = kT \frac{4(R_1 + R_2)}{(\sqrt{R_1/R_2} + \sqrt{R_2/R_1})^2} \Delta f$$

$$= 4kT \frac{R_1R_2}{R_1 + R_2} \Delta f$$

in agreement with the value obtained by applying Eq. 24 to the parallel combination of R_1 and R_2.

15. Antenna Noise. The resistance component of antenna impedance may be divided into a small part resulting from losses in the antenna and a large part, the *radiation resistance* (Art. 2, Chap. XIII), associated with the radiation properties of the antenna. As may be expected, the dissipative resistance gives rise to noise because of thermal agitation of electrons in the antenna conductors. The radiation resistance gives rise to fluctuation noise also because thermal-agitation noise radiated from the antenna surroundings is received by the antenna. Were the antenna and all its surroundings at the same temperature, Eq. 23 would specify the available antenna noise power. (To justify this statement, thermal equilibrium of the antenna and its surroundings with an antenna-load resistance is considered. Zero net power is then transferred to the resistor, and therefore the power available from the antenna equals that available from the resistor.) Of course, the objects surrounding the antenna are actually at a variety of different temperatures, and an experimentally determined effective temperature must be used to specify actual antenna fluctuation noise. Because noise power is received mainly from objects within the major lobe of the radiation pattern, the effective temperature depends upon antenna orientation.

In addition to thermal agitation of electrons, many sources of interference, such as atmospherics, ignition noise, and near-by radar-frequency equipment, contribute to the unwanted signals at the antenna output. The frequency spectrum of interference is not uniform, and at radar frequencies interference can often be neglected relative to fluctuation noise—unless other radio equipment of the same frequency is operated near by or intentional interference is present.

The fluctuation noise generated in the antenna resistance at a temperature of 290 K (17 C) is usually taken as a standard value of antenna noise. By use of this standard, the noise properties of the receiver may be separated from the noise characteristics of the antenna and wave-propagation medium. The relation of noise generated in the receiver to the

standard antenna noise is a measure of the receiver merit (see Art. 17), and laboratory noise testing may be performed with the antenna replaced by a resistor. The relation of the actual antenna noise to the standard noise is determined by the antenna directivity, the operating frequency of the radar system, and the site and orientation of the antenna. High directivity improves the signal-to-noise ratio because it reduces the response to noise and interference waves arriving from directions other than the direction to the signal source. An operating frequency in a low-interference region of the frequency spectrum is desirable.

16. Vacuum-tube and Crystal Noise.[1] Nonlinear circuit elements, like linear elements, generate fluctuation noise because of a randomness in the motion of electrons. The detailed mechanism of noise generation is different, however, and the average-square value of the noise cannot be calculated so readily. Discussion of vacuum-tube noise begins with diodes, despite the fact that these tubes are seldom used in noise-sensitive receiver-input circuits. The diode discussion aids the understanding of noise in amplifier-type tubes and in crystals. Furthermore, a special type of diode serves as a standard noise source in noise-measurement experiments.

Diodes. Noise currents are so tiny that in the absence of other varying currents a diode operates over an essentially straight portion of its nonlinear i_b, e_b characteristic. Therefore, the diode as a noise source may be represented by either of the linear equivalent circuits for a source in Fig. 20, provided the parameters in the circuit are adjusted to correspond to the operating point employed. The source impedance $Z = 1/Y$ becomes the impedance of the parallel combination of incremental plate resistance r_p and plate-cathode capacitance C_{pk}. The voltage generator E represents the open-circuit noise voltage, and the current generator I is the short-circuit noise current. *Open circuit* and *short circuit* refer to conditions in the linear equivalent circuit. A high a-c load impedance may provide essentially open-circuit conditions at frequencies of interest, even when a circuit of low d-c resistance is used to control the operating point.

The incremental resistance r_p and the average-square values of E or I in the equivalent circuit determine the available noise power of a diode. In general, the available power in a frequency range Δf is different from the linear-impedance value $kT\Delta f$ and in fact depends upon frequency and choice of operating point. The available power is, however, proportional to Δf and to the cathode temperature of the diode and therefore may be specified as $k\alpha T_k \Delta f$, where T_k is the cathode temperature and α is

[1] For a very readable summary of vacuum-tube noise see B. J. THOMPSON, "General Survey," Part I of "Fluctuation in Space-charge-limited Currents at Moderately High Frequencies," *RCA Rev.*, **4** (January, 1940), 269–285.

the *noise-temperature ratio* of the diode, or ratio of available noise power from the diode to $kT_k\Delta f$, the available noise power of a linear impedance at cathode temperature. The diode may be replaced in noise calculations by a linear resistance r_p (in parallel with C_{pk}) at a temperature αT_k. This temperature is the equivalent *noise temperature* of the diode.

The special diodes employed in noise measurements have tungsten cathodes and—in contrast to diodes in normal radio and radar circuits—are so operated that the plate current is *temperature limited.* That is, the plate voltage is made high or the cathode-heating current low so that all electrons emitted by the cathode travel to the plate. In such operation, r_p is exceedingly high because the plate current, being determined by cathode emission, is little affected by plate-voltage variations. The tube may be considered, for most purposes, a source of noise current independent of load impedance, and the concept of available power loses significance. The tube is useful as a noise standard because the average-square value of noise current can be determined with accuracy from the theoretical relation

$$\Delta|I|^2 = 2Q_eI_b\Delta f \tag{26}$$

where

$\Delta|I|^2$ is the average-square value of noise-current components in the frequency range Δf,

Q_e is the magnitude of the charge of an electron, 1.602×10^{-19} coulomb,

I_b is the average plate current of the diode.

Diode noise is a result of randomness in the emission of electrons from the cathode. Emission depends upon the thermal motions of electrons within the cathode metal, and the irregular nature of the thermal motions produces erratic fluctuations in the rate of emissions. In temperature-limited operation, the emission fluctuations appear directly as noise in the plate current and are called *shot-effect noise.* In *space-charge limited* operation, the available emission current greatly exceeds the actual plate current, and the space charge has an important smoothing effect upon plate current. The resulting noise currents have values much less than specified by Eq. 26 and are attributed to the *reduced shot effect.* Calculations[1] show the available noise power to be $0.64kT_k\Delta f$ (or noise ratio $\alpha = 0.64$) for diodes of normal geometry operated in a usual region of the space-charge-limited range and at any frequency up to a few tens of megacycles. Experiments often yield somewhat greater noise-temperature ratios because the reduced shot effect is so small that other causes of fluctuation in electron flow have a measurable effect.

[1] A. J. RACK, "Effect of Space Charge and Transit Time on the Shot Noise in Diodes," *B.S.T.J.*, **17** (October, 1938), 592–619.

Consideration of a diode with plate and cathode at the same elevated temperature in thermal equilibrium with a linear impedance might be thought a means of showing (in the same way as was done for an antenna in Art. 15) the diode and linear impedance to have equal values of available noise power and thus proving $\alpha = 1$. If noise currents alone are present in the diode-and-impedance system, such an argument is valid. Inclusion of a battery in the circuit to shift the operating point away from $e_b = 0$ provides a source of energy, destroys the thermal equilibrium, and invalidates the argument.

Crystals. Either circuit of Fig. 20 may be used to represent the noise-generating properties of a mixer crystal. Because the crystal, like the diode, is a nonlinear element (see Art. 12), the parameters of the equivalent circuit are functions of the point on the crystal characteristic about which operation occurs. The available noise power may again be expressed as $k\alpha T\Delta f$, where α is the noise-temperature ratio of the crystal.

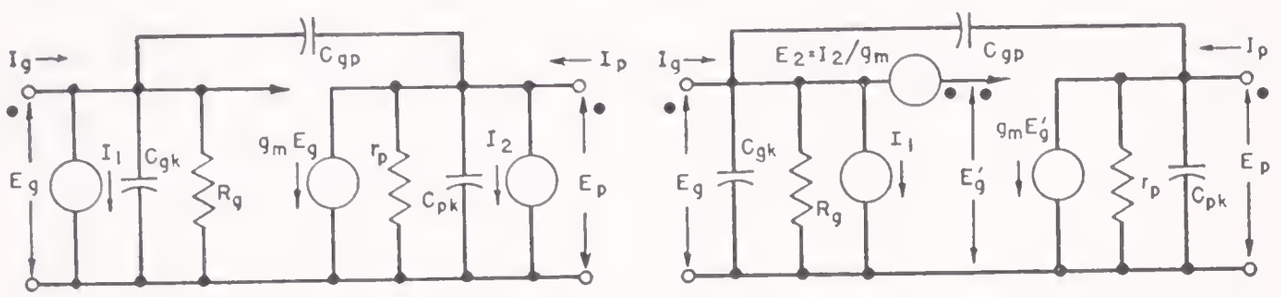
The theory of noise generation in crystals has been investigated at some length,[1] but the processes involved are not yet completely understood. Experimental values of α obtained in the range of radar intermediate frequencies (15 to 60 Mcps) range from somewhat less than unity to about 5.0 for currents in the forward direction and are several times as large for currents in the reverse direction. Theoretical calculations are in rough agreement with these values but fail to account for another experimental result: At frequencies less than about 1 Mcps, α increases in inverse proportion to the frequency and rises to 100,000 or more at low audio frequencies.

In radar-receiver mixers, the voltage at a crystal is the sum of very minute noise and signal voltages and a relatively large local-oscillator voltage. The oscillator voltage sweeps the operating point over a wide range of the crystal characteristic, and the crystal noise power of interest may be thought of as an average of the noise powers delivered during different portions of the cycle of local-oscillator voltage. The corresponding available noise power may be expressed as $ktT\Delta f$, where t is the apparent value of the noise-temperature ratio in the variable-operating-point mixer application. Values of t depend not only upon the local-oscillator power and the intermediate frequency at which the noise is measured but also to a small extent upon r-f adjustments in the crystal circuit. Typical 30-Mcps values for standard crystals are indicated by the noise-temperature ratios of Table 1 (p. 376). The significance of t is discussed further in connection with other mixer characteristics in Art. 28.

Triodes. The equivalent circuits of Fig. 20 are inadequate for representing the noise-generating properties of triodes because a triode is not a

[1] H. C. TORREY and C. A. WHITMER, *Crystal Rectifiers*, 179–197 (McGraw-Hill Book Company, Inc., New York, 1948).

two-terminal impedance. The analogue of Fig. 20*A* for an amplifier tube is drawn in Fig. 21*A*. This circuit may be considered a representation of Eqs. 20, p. 336, with the current I_1 added to the first equation and I_2 added to the second. These currents are short-circuit currents analogous to I in Fig. 20*A*. They represent noise currents at the terminals when E_g and E_p are zero, that is, when a-c short circuits are placed from grid to cathode and from plate to cathode. If *either* short circuit is removed, noise currents at *both* terminal pairs are altered, and the amount of the change can be calculated from the circuit of Fig. 21*A*. The circuit between sources I_1 and I_2 is the equivalent circuit of Fig. 14*B* and is analogous to Y in Fig. 20*A*.



(A) CURRENT-SOURCE EQUIVALENT (B) EQUIVALENT WITH I_2 REFERRED TO GRID

FIG. 21. Triode equivalent circuits including noise sources.

The magnitudes of I_1 and I_2, which depend upon the operating point of the tube and the frequencies of the noise components considered, may be determined by experiment. Approximate values can be determined from theoretical considerations. Noise arises in the space-charge-limited cathode current of a triode because of the reduced shot effect, exactly as it does in a diode. In general, cathode-current noise fluctuations divide between grid and plate circuits of the triode and contribute to both I_1 and I_2. If the grid-bias voltage is negative, and if the frequency range considered is so low that induced grid current and lead inductance may be neglected, then no noise current appears in the grid circuit; I_1 is zero, and I_2 is the result of the entire cathode-current fluctuation.

Consider the "diode" formed when an a-c short circuit is placed between grid and plate of a triode. In the frequency interval Δf, the available noise power of the diode is $k\alpha T_k \Delta f$, where α is of the order of unity (but not exactly 0.64 because of the peculiar geometry of the "diode"). In a low-frequency range, the plate conductance of the diode (determined from Fig. 21*A* with $1/R_g = 0$ and $E_g = E_p$) is $g_m + 1/r_p$, or $g_m(1 + 1/\mu)$. The average-square value of the short-circuit diode noise current is therefore, by Eq. 25,

$$\Delta|I|^2 = 4k\alpha\left(1 + \frac{1}{\mu}\right) T_k g_m \Delta f \tag{27}$$

This diode current equals the cathode current of the triode when grid and plate are both connected by a-c short circuits to the cathode; thus Eq. 27 specifies the average-square value of the sum of I_1 and I_2. Though derived from low-frequency consideration, Eq. 27 is nearly correct whenever transit times are small fractions of sine-wave periods and may be used as an approximation at radar intermediate frequencies as well as at lower frequencies.

In the same way that space-charge variations due to grid signal voltage induce components of grid current, so also noise fluctuations of space charge induce noise components of grid current. Noise entering the grid circuit in this manner is called *induced grid noise*. The magnitude of the induced noise currents can be calculated at frequencies for which the space charge at each point of the tube varies essentially in the same phase. Values so obtained are approximately correct even at frequencies as high as radar intermediate frequencies.

When the variation of space charge is everywhere in phase with i_e, the electron current leaving the cathode, the induced grid charge is $K_e i_e$, as in Art. 11, and the corresponding induced current is $K_e di_e/dt$. Because of the differentiation, each sine-wave component of induced current leads the corresponding component of i_e by 90 deg, and the amplitudes of induced-current components are ωK_e times the amplitudes of corresponding components of i_e. The *average-square* value of induced noise currents, therefore, computed from the average-square value of i_e in Eq. 27, is

$$\begin{aligned}\Delta|I_1|^2 &= (\omega K_e)^2 4k\alpha \left(1 + \frac{1}{\mu}\right) T_k g_m \Delta f \\ &= 4k\alpha K_e^2 T_k \left[\omega^2 g_m \left(1 + \frac{1}{\mu}\right)\right] \Delta f\end{aligned}$$

In this equation, the frequency interval Δf must be a small percentage of its center frequency in order that ω may be considered constant within the interval. The discussions of Arts. 10 and 11 show that $\omega^2 g_m(1 + 1/\mu)$ is proportional to $1/R_g$; that is, $\omega^2 g_m(1 + 1/\mu)$ equals K/R_g, where K is a constant of proportionality. Thus

$$\Delta|I_1|^2 = 4k(\alpha K_e^2 K) T_k \frac{1}{R_g} \Delta f$$

or

$$\Delta|I_1|^2 = 4k\alpha_i T_k \frac{1}{R_g} \Delta f \tag{28}$$

where α_i is $\alpha K_e^2 K$.

Equation 28 results because both $1/R_g$ and the average-square value of induced grid noise vary as the square of frequency. Because of this coincidence, it is convenient to think of induced grid noise as the noise in

a linear resistor R_g having a noise temperature (see Eq. 28) of $\alpha_i T_k$, or a noise-temperature ratio α_i. The value of α_i is about 1.4 if $1/R_g$ is the component of grid conductance due to transit-time effects alone.[1] Cathode-lead inductance might also be expected to cause noise in the grid circuit because of the voltage produced across the inductance by noise fluctuations of cathode current. Calculations indicate that the average-square magnitude of such noise, which increases with the fourth power of frequency, has a negligible value relative to induced grid noise even at the highest frequencies of radar i-f amplifiers.

Because the average-square value of induced grid noise is proportional to the square of frequency, I_1 may be considered zero in video-frequency amplifiers and I_2 may be taken to equal the total cathode-current noise specified by Eq. 27. In intermediate-frequency amplifiers, I_1 must be calculated from Eq. 28. Throughout the radar i-f range, the values of $\Delta|I_1|^2$ are small relative to $\Delta|I|^2$, because $1/R_g$ is small relative to g_m. [Compare Eqs. 27 and 28. The factors $\alpha(1 + 1/\mu)$ and α_i are roughly equal, as each is of the order of unity.] Therefore, the decrease of I_2 caused by cathode current diverted to the grid circuit is negligible, and Eq. 27 specifies I_2 for i-f amplifiers as well as for video amplifiers.

For convenience in noise calculations, the current generator I_2 in Fig. 21*A* is often replaced by the equivalent grid-circuit noise-voltage generator E_2 of Fig. 21*B*. The two circuits are equivalent because current source $g_m E_g'$ in Fig. 21*B* is

$$g_m E_g' = g_m(E_g + E_2) = g_m E_g + g_m \frac{I_2}{g_m} = g_m E_g + I_2$$

The change of sources is made in order that relations among shot noise, induced grid noise, and thermal noise may be visualized more readily. The *average-square* value of E_2, according to Eq. 27, is

$$\begin{aligned}\Delta|E_2|^2 &= \frac{\Delta|I_2|^2}{{g_m}^2} = 4k\alpha\left(1 + \frac{1}{\mu}\right) T_k \frac{1}{g_m}\Delta f \\ &= 4kTR_{eq}\Delta f\end{aligned}$$

where T is a standard temperature, usually taken to be 290 K, and

$$R_{eq} = \alpha\left(1 + \frac{1}{\mu}\right)\frac{T_k}{T}\frac{1}{g_m} \tag{29}$$

is the *equivalent noise resistance* of the tube. A resistor R_{eq} in the position of E_2 in Fig. 21*B* would produce the same noise in the plate circuit as does the reduced shot effect. Because $\alpha(1 + 1/\mu)$ is approximately 1 and T_k/T is about 3, R_{eq} is roughly $3/g_m$.

[1] C. J. BAKKER, "Fluctuations and Electron Inertia," *Physica*, **8** (January, 1941), 23–43.

Specification of accurate values of R_{eq} for specific tubes, operating points, and frequency ranges is a common and convenient method of stating the plate-circuit noise generation in an amplifier tube. If R_g and I_1 in Fig. 21*B* are negligible, R_{eq} has the particular advantage of showing by its relation to the input-circuit resistance the relative importance of shot noise and input-circuit thermal noise.

If induced grid current is not negligible, the phase relations among the sinusoidal components of induced-grid-noise current and equivalent shot-noise voltage must be considered in a determination of the total noise. If I_1 and E_2 are complex numbers representing single-frequency noise components, and if Z_g is the complex grid-cathode impedance, then single-frequency components of the total tube noise referred to the grid voltage are represented by $Z_g I_1 + E_2$. Current I_1 is 90 deg ahead of E_2 in phase, and $Z_g I_1$ is in quadrature with E_2 if the grid circuit is tuned so

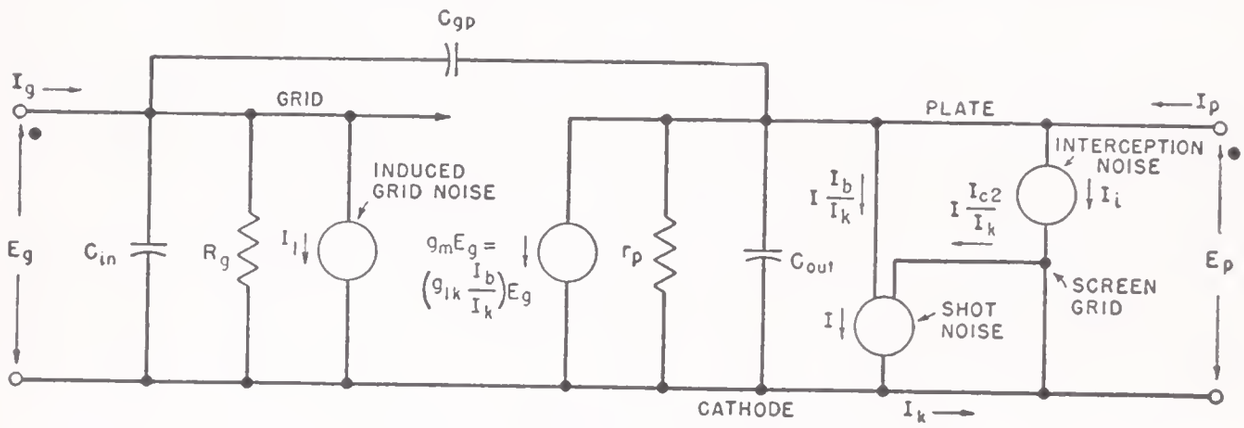


FIG. 22. Sources of noise in a pentode.

that Z_g is a resistance. Under this condition, average-square values of the two noise voltages may be added as if the two sources were independent. If the grid circuit is detuned, the quadrature phase relation is altered, and the total noise referred to grid voltage may be either greater or less than the sum of the average-square values.

Pentodes. The noise-generation equivalent circuits of Fig. 21 apply to pentodes as well as to triodes, provided the parameters in the circuit are correctly evaluated. The plate-circuit noise current I_2 of Fig. 21*A* includes noise from a source not present in triodes, which may be explained in terms of the more detailed circuit of Fig. 22. This circuit includes, for purposes of explanation, noise-current sources in the screen-grid circuit, even though such currents flow in the screen-grid-to-cathode short circuit and have no direct effect upon the amplifier output.

Current source I represents the reduced-shot-effect noise in the cathode current, which noise occurs exactly as in a diode or a triode. Source I_1 is the induced grid current caused by cathode-current fluctuations in the same way as in a triode. Variations of cathode current divide between screen grid and plate in very nearly the same proportion as does the

average cathode current. Therefore, if I_b, I_{c2}, and I_k are average values of plate, screen-grid, and cathode current, respectively, source I divides into a plate-circuit noise current $I(I_b/I_k)$ and a screen-grid-circuit noise current $I(I_{c2}/I_k)$, as indicated in the figure. Similarly, if g_{1k} is the transconductance relating cathode-current variations to voltage variations at grid 1, the grid-to-plate transconductance g_m appearing in the figure is very nearly $g_{1k}(I_b/I_k)$.

The additional noise in pentodes is called *interception noise* or *partition noise* because it results from a randomness in the interception of electrons by the screen grid. The resulting chance variations in plate current and screen-grid current are represented by the noise-current source I_i in Fig. 22. The greatest magnitude of interception noise occurs in tubes for which the random current-division process affects the entire electron stream and is[1]

$$\Delta|I_i|^2 = 2Q_e \frac{I_{c2}I_b}{I_k} \Delta f \tag{30}$$

where $\Delta|I_i|^2$ is the average-square value of the interception-noise current and Q_e is the magnitude of the charge of an electron.

Because interception noise and shot noise arise from independent sources, they are uncorrelated, and average-square values may be added to obtain plate and screen-grid noise currents. The total noise in each of these currents can never exceed the shot noise that would be present if the currents were obtained from temperature-limited cathodes. Thus interception noise has the effect of partially removing the smoothing effect of space charge.

If circuits are employed that combine noise in the screen-grid current with noise in the plate current, average-square additions cannot be made because, as indicated by the sources in Fig. 22, the shot-noise currents in the two circuits are correlated as are the interception-noise currents. For example, if an a-c short circuit is placed between plate and cathode in Fig. 22, interception noise is confined to the path from screen grid to plate and only shot noise appears in the cathode current.

For pentodes, as for triodes, the relation of tube noise to input-circuit thermal noise may be measured by means of an equivalent noise resistance. The noise currents $I(I_b/I_k)$ and I_i in the plate circuit correspond to uncorrelated grid voltages $I(I_b/I_k)/g_m$ and I_i/g_m. The average-square value of the equivalent grid voltage is therefore

$$\Delta|E_2|^2 = \frac{\Delta|I|^2}{{g_m}^2} \frac{{I_b}^2}{{I_k}^2} + \frac{\Delta|I_i|^2}{{g_m}^2}$$

[1] D. O. NORTH, "Multicollectors," Part III of "Fluctuations in Space-charge Limited Currents at Moderately High Frequencies," *RCA Rev.*, **5** (October, 1940), 244–260.

and the equivalent grid resistance is

$$R_{eq} = \frac{1}{4kT\Delta f}\left[\frac{\Delta|I|^2}{g_m{}^2}\frac{I_b{}^2}{I_k{}^2} + \frac{\Delta|I_i|^2}{g_m{}^2}\right]$$

The average-square value of the cathode-current shot noise $\Delta|I|^2$ is given by Eq. 27 if g_m in that equation is replaced by the grid-cathode transconductance g_{1k}, and the interception-noise current $\Delta|I_i|^2$ is given by Eq. 30. Thus

$$R_{eq} = \frac{1}{4kT\Delta f}\left[\frac{4k\alpha(1 + 1/\mu)T_k g_{1k}\Delta f}{g_m{}^2}\frac{I_b{}^2}{I_k{}^2} + \frac{2Q_e I_{c2}\Delta f}{g_m{}^2}\frac{I_b}{I_k}\right]$$

$$R_{eq} = \left[\alpha\left(1 + \frac{1}{\mu}\right)\frac{T_k}{T} + \frac{Q_e}{2kT}\frac{I_{c2}}{g_m}\right]\frac{I_b}{I_k}\frac{1}{g_m} \qquad (31)$$

The first of the terms within brackets in Eq. 31 has a value approximating 3, and the factor outside the brackets equals $1/g_{1k}$. Thus the component of R_{eq} due to shot noise is roughly $3/g_{1k}$, which is comparable to the equivalent noise resistance of a triode. Use of the values of Q_e and k with the assumption $T = 290$ K yields for $2kT/Q_e$ the value $\frac{1}{20}$ volt. The component of R_{eq} due to interception noise is therefore about $20(I_{c2}/g_m)/g_{1k}$, which for most tubes lies in the range $5/g_{1k}$ to $25/g_{1k}$. Interception noise may therefore be expected to exceed reduced-shot-effect noise, and pentodes in general are materially noisier than triodes.

17. Noise Figure. The over-all effect of the many noise sources in a receiver is frequently specified by means of the *noise figure* of the receiver. This quantity measures the noise generation within the receiver circuits. Antenna noise temperature measures the noise entering with the signal. The merit of the receiving system with respect to noise depends also upon the extent to which amplifiers, detector, and indicating device permit discrimination between signal and noise.

The noise figure[1] of any linear two-terminal-pair network can be defined in terms of its performance with a standard noise source connected to its input terminals. The available noise power of the standard source in a bandwidth Δf is $kT\Delta f$, with $T = 290$ K, the same as the standard value of antenna noise power (see Art. 14). The noise ΔN_o at the output of the network within the same frequency band arises from amplification of the input noise and from noise generation within the network and may be expressed as $\Delta N_{\text{ideal}} + \Delta N$, where ΔN_{ideal} is the output noise of an ideal noiseless network and ΔN is the additional noise generated in the network.

[1] D. O. NORTH, "The Absolute Sensitivity of Radio Receivers," *RCA Rev.*, **6** (January, 1942), 332–343; H. T. FRIIS, "Noise Figures of Radio Receivers," *Proc. I.R.E.*, **32** (July, 1944), 419–422.

The noise figure is defined as

$$F = \frac{\Delta N_o}{\Delta N_{\text{ideal}}} = 1 + \frac{\Delta N}{\Delta N_{\text{ideal}}} \tag{32}$$

The noise symbols in Eq. 32 may stand for available noise power at the output of the network, or (because mismatch affects both quantities of each ratio in the same way) for the noise power delivered to a load of any impedance. Alternatively, without change in the value of F, the symbols may refer to the average-square noise current or voltage delivered to any load.

Ordinarily the amplification or loss of a network is a function of frequency. Similarly, the noise per-unit bandwidth generated by elements in the network may vary with frequency. Consequently, a distinction between *single-frequency noise figure* and *integrated noise figure* must be made. If Δf signifies a bandwidth so small that variations of amplification and nonuniformity of noise spectra may be ignored, then Eq. 32 defines the single-frequency noise figure. On this basis, noise figure may be determined for various locations of the band Δf and plotted as a function of frequency. If Δf is a wide band containing all frequencies of importance in the network output, Eq. 32 defines the integrated noise figure. The noise quantity ΔN_o then becomes N_o, the total noise power (or average-square noise current or voltage) at the output; ΔN_{ideal} and ΔN become N_{ideal} (the component of ΔN_o arising from amplification of input noise) and N (the component caused by noise sources in the network).

If the noise symbols represent available power, the value of ΔN_{ideal} for a narrow bandwidth Δf may be expressed as $GkT\Delta f$, $kT\Delta f$ being the available noise power of the source and G the *available-power gain* of the network at the frequency of the band considered. The single-frequency noise figure then becomes

$$F = \frac{\Delta N_o}{GkT\Delta f} = 1 + \frac{\Delta N}{GkT\Delta f} \tag{33}$$

The total available noise power resulting from amplification of the input noise is the sum of the contributions of many small frequency bands that together cover the entire frequency range. This sum may be expressed as the integral $\int_0^\infty GkTdf$, and the integrated noise figure is therefore

$$F_{\text{int}} = \frac{N_o}{\int_0^\infty GkTdf}$$

N_o being the total noise power available from the network output terminals.

Another common form for the single-frequency noise figure is obtained if the source is supposed to supply, within a narrow frequency band, both the noise ΔN_i and a signal ΔS_i, the resulting output signal being ΔS_o. These symbols, like those of Eq. 32, may stand for available power, delivered power, or average-square current or voltage, provided only one interpretation is used throughout a relationship. Within the narrow band considered, $\Delta S_o/\Delta S_i$ equals $\Delta N_{\text{ideal}}/\Delta N_i$, because each ratio is the gain of the network. Therefore, from Eq. 32,

$$F = \frac{\Delta N_o}{\Delta N_{\text{ideal}}} = \frac{\Delta N_o}{(\Delta S_o/\Delta S_i)\Delta N_i} = \frac{\Delta S_i/\Delta N_i}{\Delta S_o/\Delta N_o} \tag{34}$$

Equation 34 states that the noise figure is the ratio of the input signal-to-noise ratio to the output signal-to-noise ratio.

Noise figures of networks in cascade (consecutive stages of a receiver, for example) can be combined in a simple manner. As an illustration of the method, consider a cascade of three networks. Let G_1, G_2, and G_3 be the available-power gains of the first, second, and third networks, respectively, and let F_1, F_2, and F_3 be the corresponding noise figures. The over-all available-power gain is $G = G_1G_2G_3$. The over-all noise figure in terms of Eq. 33 is

$$F = 1 + \frac{\Delta N}{GkT\Delta f} = 1 + \frac{\Delta N_1G_2G_3 + \Delta N_2G_3 + \Delta N_3}{G_1G_2G_3kT\Delta f}$$

where ΔN is the available output noise power due to noise sources in all three networks, ΔN_1 is the available noise power at the output of the first network caused by noise sources within that network, and ΔN_2 and ΔN_3 are corresponding quantities for the second and third networks. Thus

$$F = 1 + \frac{\Delta N_1}{G_1kT\Delta f} + \frac{1}{G_1}\frac{\Delta N_2}{G_2kT\Delta f} + \frac{1}{G_1G_2}\frac{\Delta N_3}{G_3kT\Delta f}$$

or

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1G_2} \tag{35}$$

Equation 35 indicates that the effect of noise sources in stages of a receiver other than the first is reduced because of the division by the available-power gains of all preceding stages. As a result, the noise figure of a receiver usually equals the noise figure of the combination of r-f amplifier (if any), mixer, and first i-f amplifier. Observe, however, that, if the available-power gain of the first stage is less than 1, the effect of noise generation in the second stage is exaggerated. Therefore, every effort is made to minimize attenuation of signals in input circuits. Such attenuation may occur in r-f systems, in mixers, and in i-f connecting cables.

Noise figure and available-power gain differ from many quantities used to describe a network in that they depend not only upon the network itself but also upon the impedance of the input source. Available-power gain, for example, is a maximum when the source impedance is the complex conjugate of the network input impedance. Output-impedance relations have no direct effect upon noise figure and available-power gain.

An example of the effect of circuit adjustments on noise figure is furnished by the tuned amplifier-input circuit of Fig. 23*A*. Current generator I is the source of signal, and resistor R (assumed to have a noise temperature of 290 K) is a source of noise entering with the signal. Resistors R_g and R_{eq} are additional noise sources that cause the noise figure to exceed unity. Resistor R_{eq} represents shot noise and interception noise generated in the plate circuit of the tube and is assigned the noise temperature 290 K. Resistor R_g is the parallel combination of

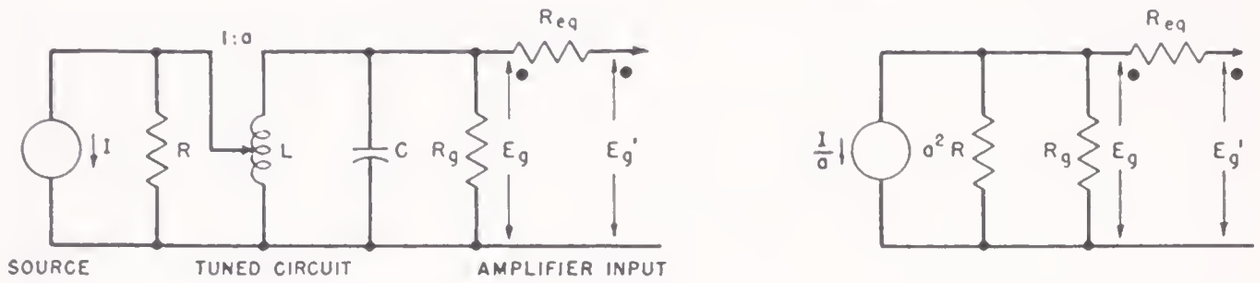


FIG. 23. Amplifier input circuit illustrating noise-figure relationships.

grid-cathode circuit resistance (including the effect of tuned-circuit losses) and the grid-loading resistance of the tube. If induced grid noise is important, the noise temperature of R_g may be high.

The single-frequency noise figure of the input circuit is to be evaluated at the resonant frequency of L and C. Within a narrow band surrounding this frequency, the tapped-inductance resonant circuit is equivalent to an autotransformer of step-up ratio a (see Art. 11, Chap. IX), and the simplified equivalent circuit of Fig. 23*B* applies. Very small values of a make the transformed source resistance a^2R small relative to R_g and the transformed average-square noise current of the source large relative to that of R_g. Thus, if R_{eq} is ignored, the noise figure approaches unity for small values of a for which the source is connected essentially to an open circuit. Small values of a, however, make the voltage E_g (both signal and noise components) very small and increase the importance of the noise of R_{eq}. With respect to R_{eq} alone, the optimum adjustment of a is the matched-resistances adjustment $a^2R = R_g$ that provides the largest possible voltage E_g. A compromise value of a between zero and the value corresponding to a resistance match provides the smallest combined effect of the two noise sources and results in a minimum noise figure.

The minimum noise figure attainable in an amplifier depends upon the center frequency and the bandwidth requirements. In a low-frequency, narrow-band amplifier, R_g can be a megohm or more. Thus a can be adjusted so that a^2R is small relative to R_g, yet a^2R and R_g in parallel are large relative to R_{eq} (for low-noise tubes, R_{eq} may be less than 1,000 ohms). In this way both the noise source associated with R_g and the noise source represented by R_{eq} are made to have negligible importance, and the noise figure is reduced nearly to unity. In high-frequency amplifiers, R_g may be little larger than R_{eq}, and the noise temperature of R_g may be high. Noise figures appreciably greater than unity are thus unavoidable, and the merit of a tube depends upon the relation of R_g to R_{eq}.

Radar-receiver noise figures are limited in large part by the low ratio of R_g to R_{eq} unavoidably present in high-frequency amplifiers. The r-f amplifiers sometimes employed in uhf receivers have moderately high noise figures for this reason. The i-f amplifiers of microwave systems have noise figures somewhat greater than unity, and the over-all receiver noise figure is considerably increased because of the less-than-unity gain of the mixer and because of mixer noise. Noise figures are often quoted in decibels. If F is a noise figure expressed as a ratio, then F_{db}, the noise figure in decibels, is

$$F_{db} = 10 \log F \tag{36}$$

Noise figures of radar receivers operating in several frequency bands are listed in Table 2. These figures are values for typical receivers, not the best obtainable values.

TABLE 2. TYPICAL NOISE FIGURES OF RADAR RECEIVERS

Frequency band of system, Mcps	*Noise figure*	
	Ratio	*Decibels*
200	6	8
1,000	10	10
3,000	13	11
9,000	25	14
24,000	40	16

18. Receiver Frequencies and Bandwidths. The responses of the r-f, i-f, and video-frequency sections of a receiver (see Fig. 1, p. 339) to low-amplitude sine waves of variable frequency are shown by the curves of Fig. 24*A*. The curve of r-f response represents the ratio of r-f mixer voltage to the voltage of a signal generator substituted for the antenna, the internal impedance of the generator being equal to the antenna

impedance. The i-f response is the ratio of the detector-input voltage to the intermediate-frequency voltage obtained from the mixer, and the video circuit response is the ratio of the voltage applied to the indicator tube to that obtained from the detector.

The center frequency of the r-f curve is f_{rf}, the nominal radio frequency of the system. Broad-band r-f circuits are generally employed to allow

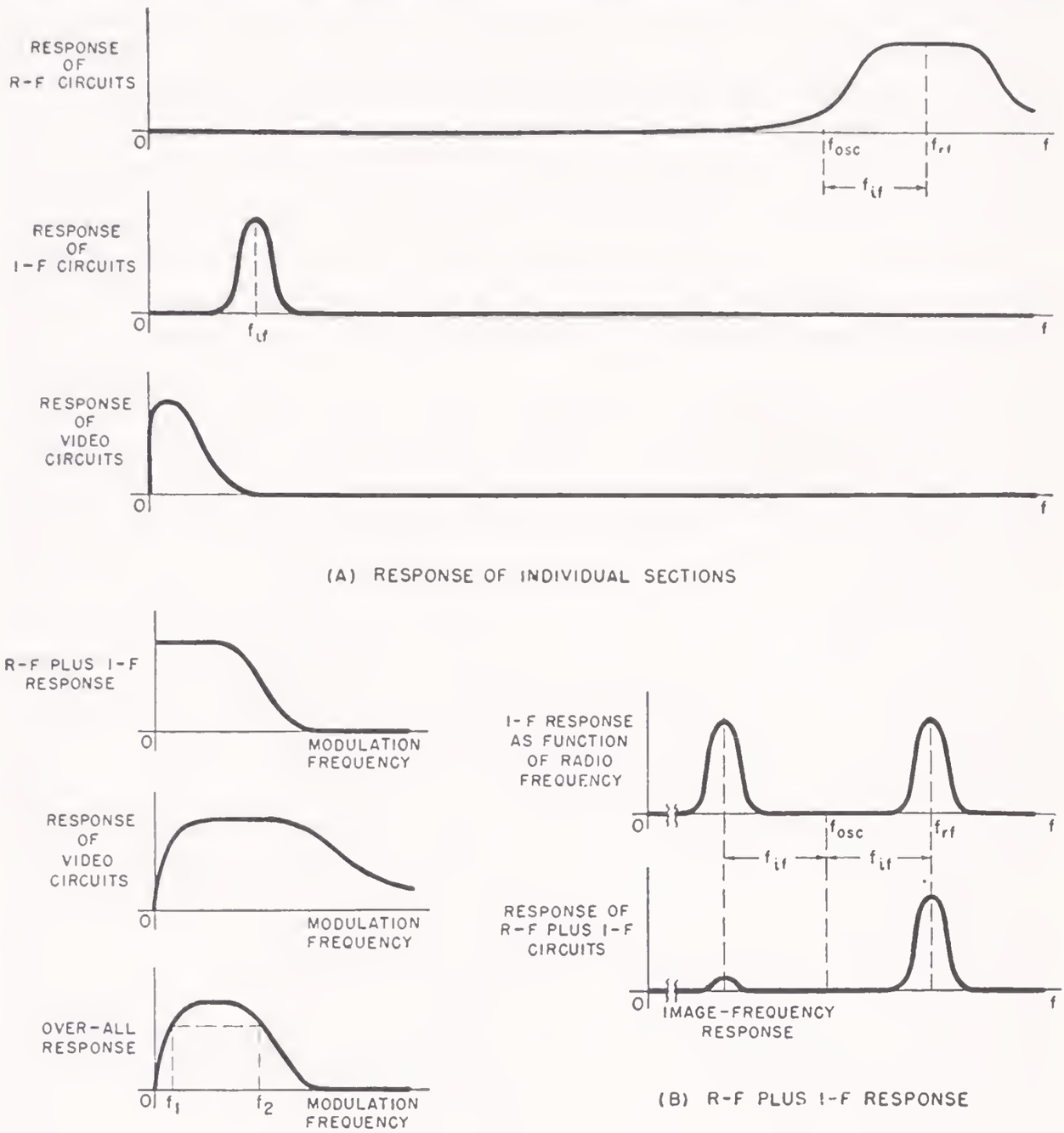


FIG. 24. Amplitude-frequency response curves of receiver.

for normal variations of transmitter frequency. The i-f curve, centered at f_{if}, is the composite response of a number of tuned stages and is narrower than the r-f curve. The i-f amplifier provides most of the bandwidth limitation of the receiver as well as the major portion of the gain. Standard radar intermediate frequencies used in this country are 15, 30, and 60 Mcps.

The r-f and i-f response curves are combined in Fig. 24*B*. An r-f input voltage of a frequency either above or below the local-oscillator frequency f_{osc} gives rise to a difference frequency that may fall within the response band of the i-f amplifier. Therefore, the response of the i-f amplifier, when plotted as a function of the input frequency that causes the i-f signal, has the double-peaked shape indicated in the first part of Fig. 24*B*. The product of this curve and the r-f response curve is the response of the r-f and i-f circuits combined and has the shape shown in the figure. The large peak in the combined curve represents the desired response of the receiver. The small peak is called the *image-frequency* response.

The curve of video-circuit response cannot be combined with the r-f and i-f curves in a similar manner because a single sine wave applied at the receiver input can cause only a steady direct voltage at the detector output, and no video-circuit output results. To measure over-all characteristics, a sinusoidally modulated input wave of carrier frequency f_{rf} is employed. The modulation frequency is low relative to f_{if}, and therefore the carrier and side frequencies that make up the modulated wave (see Art. 4) fall within the frequency range of the large peak of the r-f-plus-i-f curve. This peak is approximately symmetrical about f_{rf}, and thus the effect of r-f and i-f circuits upon the envelope of the input signal is the same as the effect of a video circuit having the response indicated by the first curve of Fig. 24*C*, which is one half of the r-f-plus-i-f response peak plotted against the frequency deviation from f_{rf}. The over-all response of the receiver to sinusoidally modulated waves of carrier frequency f_{rf} is therefore the product of the r-f-plus-i-f and video curves in Fig. 24*C*, and this product is shown by the final curve in that figure. The frequency scales in Fig. 24*C* are expanded relative to the scales of Figs. 24*A* and *B*.

The response of the receiver to echo pulses of carrier frequency f_{rf} may be determined from the last curve in Fig. 24*C* by the method explained in Art. 7. The lower half-power frequency f_1 (which depends on the video circuits) determines the maximum length of pulse that can be reproduced. If the pulse length exceeds about $1/f_1$, appreciable peaking and overshoot may be expected. The upper half-power frequency f_2 (which depends upon r-f, i-f, and video circuits) determines the steepness of the edges of the reproduced pulse. For pulse durations less than about $1/f_2$, a decrease of amplitude as well as a loss of edge sharpness occurs.

Although bandwidth limitations in the r-f, i-f, and video circuits have equivalent effects upon the reproduction of pulse edges, they are not equivalent in some other respects. For example, r-f and i-f characteristics alone determine the relative response of the receiver to signals of different carrier frequency. If the transmitter frequency is somewhat variable, wide-band r-f circuits are necessary if the receiver is always to be in tune. Wide-band i-f circuits also are required unless manual or auto-

matic tuning is used to compensate for the changes of transmitter frequency.

With respect to noise also, r-f, i-f, and video bandwidths have somewhat different effects. Because fluctuation noise from all sources has an average-square value proportional to bandwidth (see Arts. 13, 14, and 15), there is in general an advantage in making the bandwidth of each section no wider than is necessary for signal reproduction. Radio-frequency and i-f selectivity have equivalent effects on noise entering with the signal, except that noise in the image-frequency band can be reduced only by control of the r-f response. In microwave receivers, noise generated by the first i-f stage is usually of major importance. Such noise is, of course, independent of r-f bandwidth. The effect of video bandwidth with regard to noise is somewhat different from that of r-f and i-f bandwidths because of a special characteristic of the detector. If the rms noise amplitude at the detector equals or exceeds the signal amplitude, the video-signal output is reduced by the presence of the noise (see Art. 26), and the signal-to-noise ratio at the detector output is less than the signal-to-noise ratio at the detector input. Bandwidth limitation in the i-f amplifier reduces the noise amplitude at the detector and minimizes suppression of the signal by the noise. From the point of view of signal discernibility and noise, therefore, the major bandwidth restriction of the receiver should occur in the i-f amplifier.

The bandwidth of each receiver section and the center frequency of the i-f section are to a considerable degree subject to control in the design of a receiver, and values are selected with a view to obtaining optimum receiver performance. Consider a particular case—selection of the i-f bandwidth to maximize the ratio of peak echo-pulse voltage to rms noise voltage at the detector.

The variation of signal and noise voltages with bandwidth is indicated in Fig. 25. Rms noise voltage, represented by the dashed curve, varies as the square root of bandwidth because average-square noise voltage is proportional to bandwidth. Peak signal voltage, to which the solid curve applies, is proportional to bandwidth when the bandwidth is narrow and approaches a constant value for wide bandwidths. This type of variation occurs because the number of components utilized in the pulse spectrum of Fig. 5*D*, p. 347, is proportional to bandwidth. For narrow bandwidths, the components are essentially equal in magnitude and at the peak of the pulse add to a sum proportional to their number. For wider bands, the magnitude of the added components is less, so that pulse height grows less rapidly with bandwidth. For bandwidths greater than $2/T$ (where T is the pulse duration), components of reversed polarity are added and a double-peaked pulse envelope is formed, as in Fig. 4*C*. Because increase of bandwidth causes a more rapid increase of signal than

noise in the narrow-band range and a less rapid increase of signal in the wide-band range, maximum signal-to-noise ratio is obtained for an intermediate value of bandwidth, which according to Fig. 25 is somewhat less than $2/T$.

The curves of Fig. 25 indicate in a general way the bandwidth required for best detection of small signals amid fluctuation noise. The exact optimum i-f bandwidth generally lies within the range $1/T$ to $2/T$, and large deviations from the optimum value cause only a minor loss in signal discernibility. The requirement for maintaining the receiver in tune with the transmitter must also be considered, and therefore a tendency to increase the bandwidth exists. Intermediate-frequency bandwidths of about $2/T$ are typical in receivers intended primarily to detect objects at

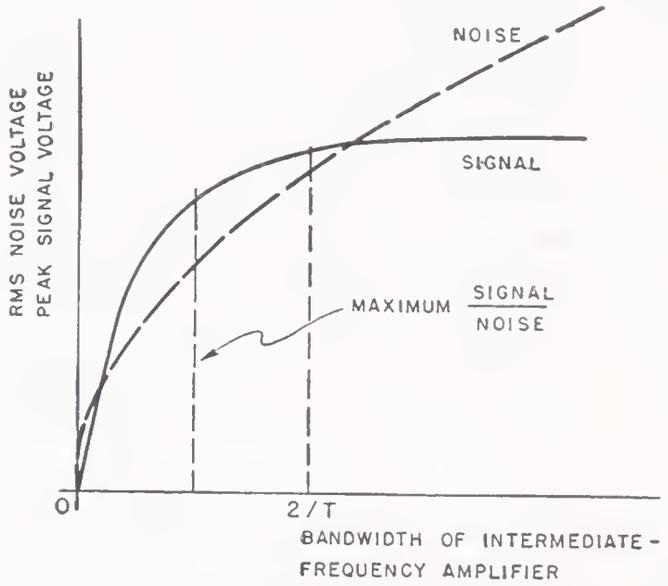


FIG. 25. Relation of rms noise voltage and peak signal voltage to intermediate-frequency bandwidth.

great range. If range accuracy and resolution are more important than maximum range, wider bandwidths are employed to increase pulse-edge steepness. For a pulse duration of 1 μs, an i-f bandwidth of 2 Mcps, a much greater r-f bandwidth, and a video bandwidth of several megacycles might be expected in a search-type receiver. For accurate ranging, the i-f bandwidth might be increased to 5 Mcps. Technical difficulties of receiver construction increase with increasing bandwidth. Therefore, bandwidths less than optimum are sometimes employed, especially if the optimum value exceeds 5 Mcps.

An intermediate frequency of one of the standard values, 15, 30, and 60 Mcps, is ordinarily chosen. The lower values have a considerable advantage with respect to noise: Grid loading and induced grid noise are reduced, and the noise figure of the first i-f stage is brought closer to unity. Reduced i-f noise is especially important in microwave systems which employ no r-f amplifier. Another advantage of lower intermediate frequency is the less serious misalignment of the amplifier stages caused

by small changes in tube and circuit capacitances. Excessively low values of intermediate frequency, on the other hand, make difficult the separation of intermediate-frequency ripple from the video frequencies at the output of the detector. Low intermediate frequencies also reduce the separation of oscillator frequency, image frequency, and desired response frequency on the r-f scale. As a result, the noise contributed by image-frequency response is increased. More important, the klystron local oscillators of microwave receivers generate appreciable noise in a band near the oscillation frequency, and the portion of this noise entering the receiver is increased. Special mixer circuits in which oscillator noise is canceled are available (see Art. 17, Chap. XII).

A 30-Mcps intermediate frequency may be considered typical of a 3,000-Mcps system having a 1-μs pulse. For higher radar frequencies or shorter pulse durations a 60-Mcps i-f amplifier may be advantageous. Ultrahigh-frequency receivers, on the other hand, generally employ a pulse duration in excess of 1 μs and have a 15-Mcps intermediate frequency.

B. COMPONENT CIRCUITS

The principles developed in Sec. A are applied in this section to the components of the block diagrams of Fig. 1, p. 339. The order of presentation—video amplifiers, i-f amplifiers, detectors, mixers, afc circuits, and anti-interference circuits—is chosen to permit logical development of the subject matter and is not at all the order in which signals traverse the receiver.

19. Typical Video Amplifier. One of the many possible connections of video circuits in a receiver is shown in Fig. 26. The video amplifier receives negative echo pulses from the detector and supplies larger amplitude pulses to a Type A indicator and to a plan-position indicator. The Type A indicator is assumed to be housed in the same chassis as the video amplifier, whereas the plan-position indicator is located in an adjacent unit and is connected to the video amplifier by means of a short cable.

Negative pulses from the detector applied to the grid of pentode V_1 produce positive plate-voltage pulses, which are applied through the cathode follower V_3 and the interunit cable to the grid of the intensity-modulated indicator tube V_7. Positive pulses from the cathode follower are also amplified by triode V_4. The negative output pulses are applied to the bottom deflection plate of the Type A indicator, thus causing an upward deflection of the electron beam. Positive range-marker pulses applied to the cathode of V_1 have the same effect as negative grid-voltage pulses and provide comparison marks for range measurement with either indicator.

To prevent the interunit-cable capacitance from reducing the high-frequency response of tube V_3 and thus distorting the pulse edges, the cable is terminated in its characteristic resistance. As a result, the input resistance of the cable equals the characteristic resistance and is very low—100 ohms, for example. Tube V_3 is therefore required to operate into a low-impedance load, and the cathode-follower connection is used for this reason. With a load impedance of 100 ohms or less, only slightly more amplification could be obtained from an amplifier than from a cathode follower, and the coupling capacitor required between amplifier plate terminal and the low-resistance cable would be large and expensive and have considerable capacitance to ground.

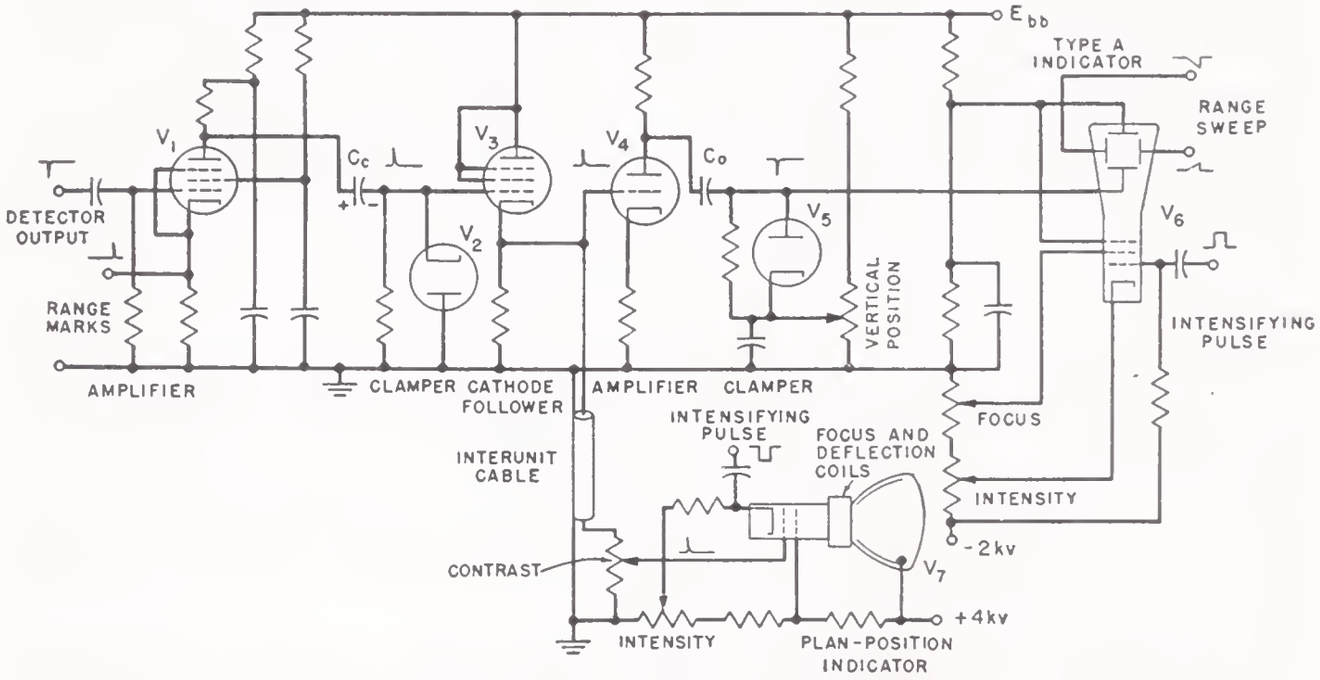


FIG. 26. Typical video amplifier.

In order that strong signals may not produce excessive brilliance and halation at the screen of the plan-position indicator, the circuit is designed to limit the maximum amplitude of the indicator grid-voltage pulses. In Fig. 26 this limiting occurs because of plate-current cutoff in tube V_1. The process is illustrated by Fig. 27, in which the full line represents actual plate voltage and the dashed line is the voltage that would result if V_1 were linear over a wide range. The highest peaks of the dashed curve correspond to a detector output voltage of roughly 25 volts, whereas tube V_1 is cut off by negative pulses of about 2-volt amplitude. Therefore, clipping or limiting of the output signal, referred to as *video-amplifier saturation*, occurs. Saturation signals produce 10 to 20 volts across the contrast-control resistor, and the adjustable tap on this resistor is set to provide the desired change of indicator-spot brilliance between no-signal and saturation-signal conditions. Optimum intensity relations in the plan-position display are obtained by the correct combination of adjustments of contrast, intensity, and receiver-gain controls.

The limiting effect in tube V_1 is important also in connection with *blocking* of the receiver. Suppose that the positive pulses available from tube V_1 were of such large amplitude that the grid-to-cathode voltage of V_3 became positive during the peak of the pulse. The resulting grid current would place an excess charge on capacitor C_c in the direction indicated by the plus and minus signs in Fig. 26. Immediately after the pulse, the additional charge on C_c would cause a negative voltage at the grid of V_3, possibly of large enough magnitude to cut off V_3. Thus little or no amplification of weak signals occurring after strong signals would be obtained. The video amplifier is said to be blocked while the excess charge remains on C_c.

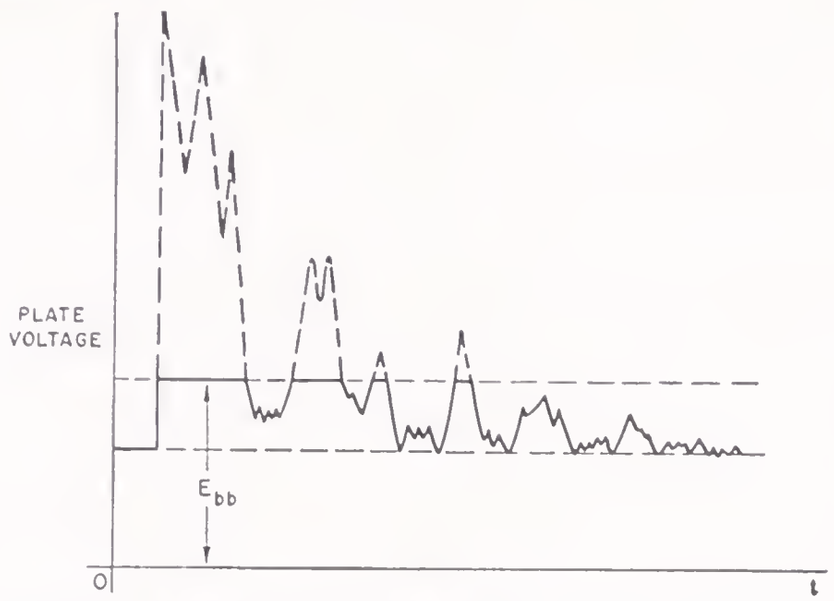


FIG. 27. Voltage waveform at plate of tube V_1 in Fig. 26.

One method of preventing video-amplifier blocking, illustrated by Fig. 26, is to use a detector that provides negative video pulses so that blocking of the first video-amplifier stage is impossible and to design the first stage so that its output pulses cannot be great enough to cause grid current in the second stage. Use of a cathode follower in the second stage (V_3 in Fig. 26) makes such a design easier to obtain. If the detector provides positive output pulses, direct coupling is often employed between the detector and the first video amplifier. The negative pulses from the first amplifier cannot cause the second stage to block, and the second-stage output is limited to prevent blocking of later stages.

Clamper tube V_2 in Fig. 26 is helpful if blocking charges do occur on capacitor C_c, because any negative voltage at the grid of V_3 causes V_2 to conduct. Thus C_c is discharged rapidly and the duration of the blocking interval is minimized. Even when limiting in V_1 prevents grid current in V_3, an excess charge is accumulated by C_c whenever a saturation signal of long duration occurs, and tube V_2 serves to remove this charge quickly at the end of the signal. Capacitor C_c could be made large to reduce the

voltage change caused by the charge accumulated, but then, without the clamper, the average value of the grid-to-ground voltage for V_3 would be zero, and the voltage between pulses would have a negative value varying with the number and strength of the echo pulses present. The between-pulse intensity level of the spot on indicator tube V_7 would therefore be variable. Use of tube V_2 ensures a fixed no-signal grid voltage for the plan-position indicator, from which value the grid voltage rises when echo pulses occur. A similar variation of the base line of the Type A display could occur because of coupling capacitor C_c, and clamper tube V_5 serves to fix the base line of this display.

20. Video-pulse Amplification. In order to understand in more detail the amplification of pulses by circuits such as those in Fig. 26, consider again the first stage of Fig. 26 which is redrawn in Fig. 28 with assumed

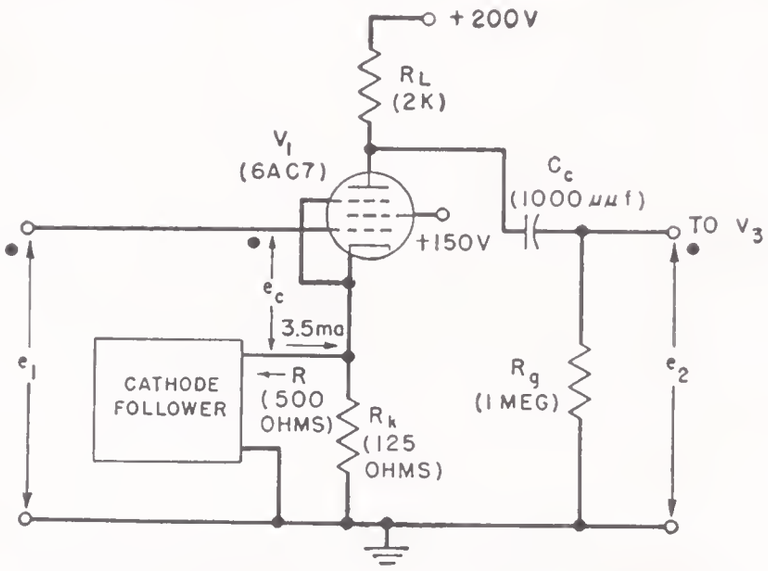


FIG. 28. First stage of video amplifier of Fig. 26.

parameter values marked. The 200-volt plate supply and 150-volt screen-grid supply indicated are obtained from the R-C circuits connected to the main supply marked E_{bb} in Fig. 26. The range-marker pulses are assumed to be derived from a cathode follower which, when no marker pulses are present, supplies a current of 3.5 ma to the cathode terminal of V_1 and which has an output resistance R of 500 ohms. The clamper tube V_2 is omitted from Fig. 28 to simplify the calculations; its effect is considered later.

The quiescent operating conditions of the tube in Fig. 28 can be obtained by trial-and-error calculations from the plate characteristics of the Type 6AC7 pentode. The plate current so obtained is 9.8 ma, and the screen-grid current is 2.7 ma. The current through resistor R_k is the sum of these two currents plus 3.5 ma from the cathode follower, or 16 ma. The voltage across R_k is $125 \times 0.016 = 2$ volts, or the quiescent grid-bias voltage is -2 volts. This value, together with the specified plate and screen-grid supply voltages, results in the electrode currents stated. At

this operating point the transconductance of the 6AC7 may be taken to be 9,000 μmhos, and the plate resistance is approximately 0.6 megohm.

The quiescent voltage across the 2-K load resistor R_L is $2 \times 9.8 = 19.6$ volts. Negative pulses at the grid of tube V_1 can, at most, cut off plate current and reduce the load-resistor voltage to zero. Thus 19.8 volts is the maximum amplitude of the positive pulses that may be delivered to the grid of tube V_3. If V_3 is a 6AG7 connected as a triode with a 300-volt plate supply and a 100-ohm cathode load resistance, a grid-to-ground voltage of about 22 volts is required to cause grid current. Because this

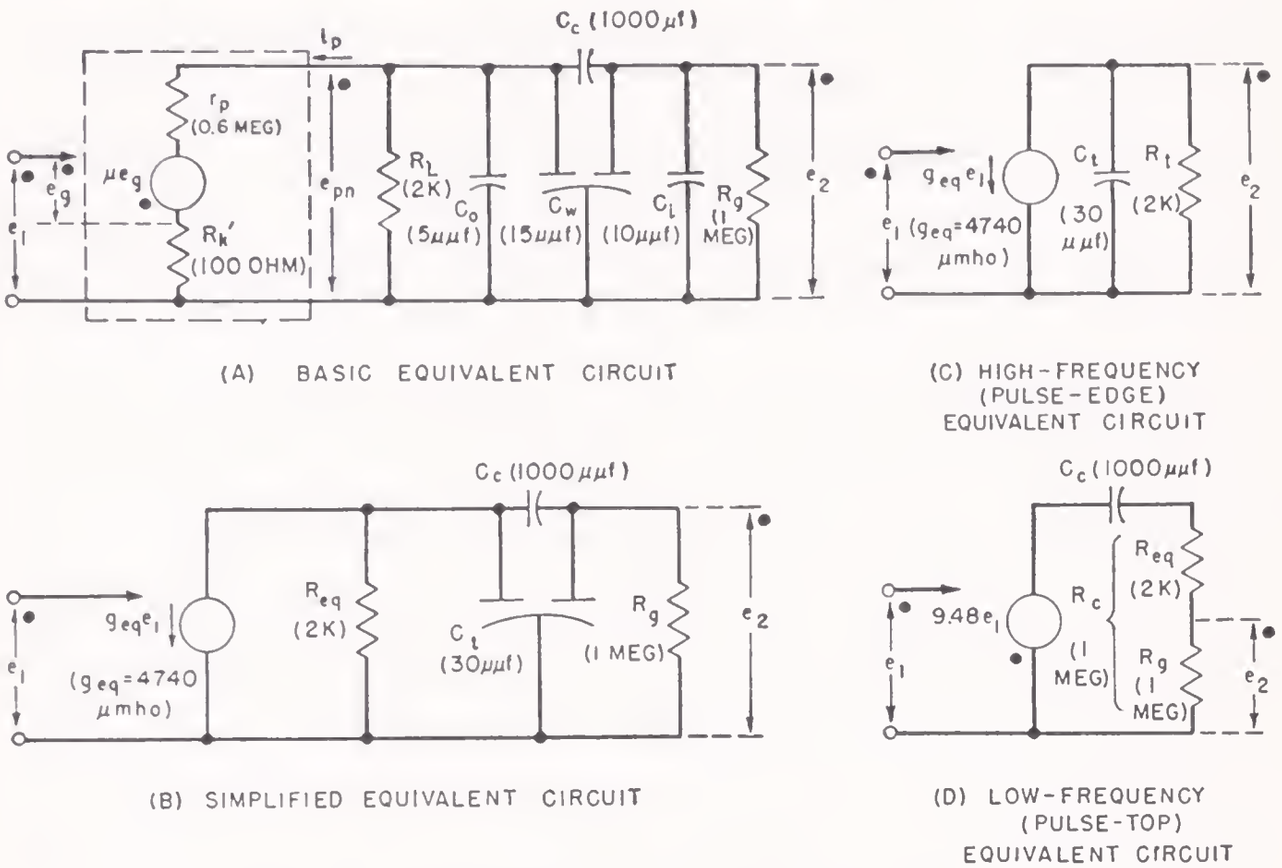


FIG. 29. Equivalent circuits for amplifier of Fig. 28.

value exceeds the 19.8 volts available, strong signals are prevented from blocking the video amplifier.

If an input voltage e_1 of small amplitude is applied to the amplifier of Fig. 28, so that tube V_1 operates within a linear region near its quiescent point, the equivalent circuits of Fig. 29 are applicable. In Fig. 29*A*, V_1 is represented as the series combination of voltage source μe_g and plate resistance r_p (e_g is the deviation of the total grid voltage e_c from its quiescent value). Resistor R_k', the resistance between the cathode of V_1 and ground, is the parallel combination of resistances R_k and R in Fig. 28, or

$$\frac{125 \times 500}{125 + 500} = 100 \text{ ohms}$$

Capacitor C_o is the output capacitance of tube V_1, C_i is the input capaci-

tance of the cathode follower V_3, and C_w is the capacitance to ground of the interstage wiring and coupling capacitance.

A simplification of the equivalent circuit is presented in Fig. 29*B*. The capacitances C_o, C_w, and C_i are added and the sum is called C_t, the total capacitance to ground. The portion of Fig. 29*A* within the dashed rectangle is replaced by an equivalent parallel combination of current source and resistance. The equivalence is established from the relation of i_p (see Fig. 29*A*) to e_{pn}, e_g, and e_1. Thus from Fig. 29*A*

$$i_p(r_p + R_k') = \mu e_g + e_{pn}$$

and

$$e_g = e_1 - R_k' i_p$$

or

$$i_p(r_p + R_k') = \mu e_1 - \mu R_k' i_p + e_{pn}$$

and

$$i_p = g_{eq} e_1 + \frac{1}{r_p + R_k'(1 + \mu)} e_{pn} \tag{37}$$

where g_{eq} is equal to $1/[1/g_m + R_k'(1 + \mu)/\mu]$. Equation 37 indicates that i_p is the sum of a current $g_{eq}e_1$ and the current that the plate-to-ground voltage e_{pn} would send through a resistance $r_p + R_k'(1 + \mu)$. Thus the circuit within the dashed rectangle is equivalent to a current source $g_{eq}e_1$ in parallel with a resistor $r_p + R_k'(1 + \mu)$. The value of $\mu = g_m r_p$ is 5,400. Thus $r_p + R_k'(1 + \mu)$ is 1.14 megohms, and this value of resistance in parallel with R_L (2 K) yields essentially 2 K, the value of R_{eq} in Fig. 29*B*. The equivalent transconductance g_{eq} is

$$\frac{1}{\frac{10^6}{9{,}000} + 100} \text{ mho}$$

or 4,740 μmhos, as indicated.

Had there been no resistance in the cathode circuit of V_1, Fig. 29*B* could have been drawn directly. Instead of g_{eq}, the transconductance g_m would have appeared, and R_{eq} would have been the parallel combination of R_L and r_p (essentially equal to R_L because of the large magnitude of r_p). The effect of the cathode resistor, therefore, is merely to reduce the amplification in the ratio g_{eq}/g_m, which is 0.526.

The circuit of Fig. 29*B* can be further simplified if the high-frequency and low-frequency performances of the circuit are considered separately. At high frequencies, capacitor C_c is effectively a short circuit, so that Fig. 29*B* reduces to Fig. 29*C*, R_t being the parallel combination of R_{eq} and R_g. (Because of the high value of R_g, R_t does not differ significantly from R_{eq}.) Similarly, at low frequencies, the current in C_t is negligible, and in Fig. 29*D* this capacitor is omitted. In this figure the parallel combination

of $g_{eq}e_1$ and R_{eq} is replaced (in accordance with Thévenin's theorem) by a voltage source $g_{eq}R_{eq}e_1 = 9.48e_1$ in series with R_{eq}, and the positions of R_{eq} and C_c are interchanged.

The circuit of Fig. 29*C* is like that of Fig. 8*A*, p. 355. Thus the response of the amplifier corresponds to Fig. 8*C* and falls off at high frequencies. The time constant of the circuit of Fig. 29*C* is R_tC_t, and the upper half-power frequency of the amplifier is thus $1/2\pi R_tC_t$. In a similar manner, the circuit of Fig. 29*D* corresponds to the one in Fig. 6*A*, p. 350, and consequently the amplifier response falls off at low frequencies in the manner of the curve of Fig. 6*C*. The lower half-power frequency is $1/2\pi R_cC_c$ when R_c is $R_{eq} + R_g$, as indicated in Fig. 29*D*. Observe that:

1. *The upper half-power frequency is determined by C_t, the total shunt capacitance, and R_t, the resistance this capacitance faces in the equivalent circuit.*

2. *The lower half-power frequency is determined by C_c, the coupling capacitor, and R_c, the resistance this capacitor faces in the equivalent circuit.*

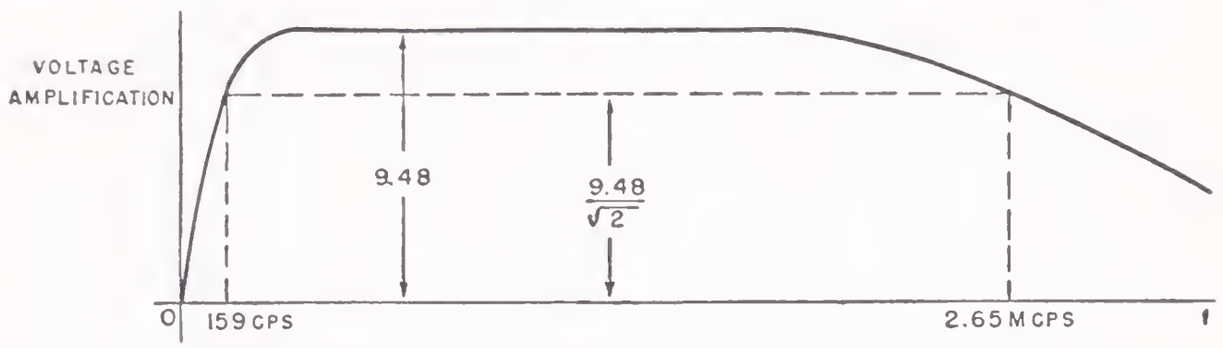


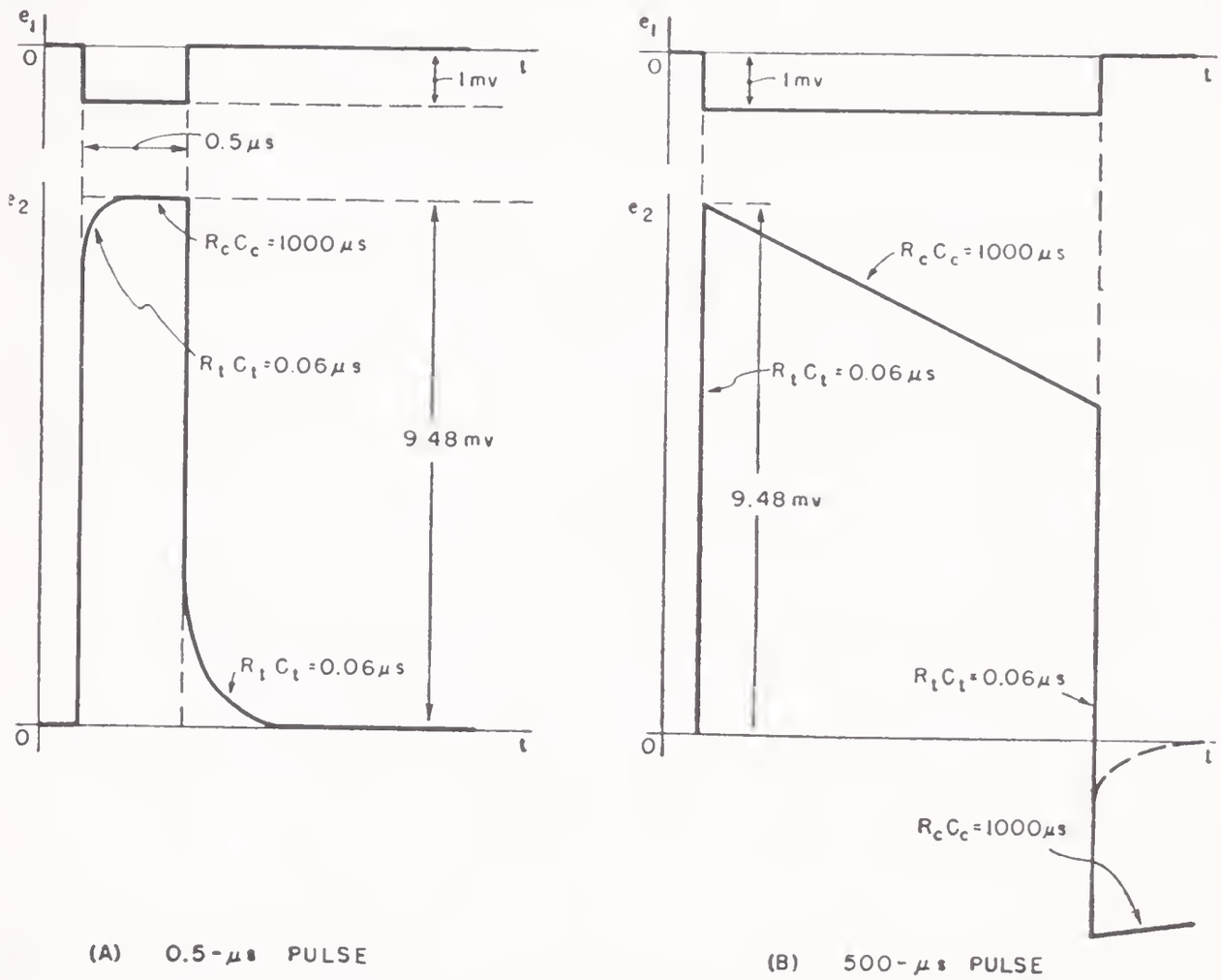
FIG. 30. Frequency characteristic for amplifier of Fig. 28.

For the circuit of Fig. 29, the amplitude response as a function of frequency is plotted in Fig. 30. The value of R_tC_t is 0.06 μs, and the upper half-power frequency is thus $1/2\pi \times 0.06$, or 2.65 Mcps. The value of R_cC_c is 1.000 μs, or 0.001 sec, and the lower half-power frequency is $1/2\pi \times 0.001$, or 159 cps. The mid-band amplification, according to Fig. 29*C*, is $g_{eq}R_t$, which is 4.74×2, or 9.48. The same value is obtainable from Fig. 29*D*.

Figure 29*C*, the high-frequency equivalent circuit, may be regarded also as an equivalent circuit that determines response to the edge of a pulse—capacitor C_c in Fig. 29*B* cannot charge or discharge appreciably during the brief time interval at a pulse edge. Similarly, because capacitor C_t draws little current during the periods between pulse edges, Fig. 29*D* is an equivalent circuit for determining response during these periods. The pulse response of the amplifier is thus a combination of the responses indicated in Figs. 9 and 10, pp. 357 and 358. The effect of the amplifier on two rectangular 1-mv pulses is thus as shown in Fig. 31. The first pulse is 0.5 μs in duration and is noticeably distorted only because of the limited high-frequency response of the circuit. The

second pulse lasts for 500 μs, and the only apparent distortion (slope of top and overshoot) is produced by lack of low-frequency response.

Clamper tube V_2 of Fig. 26, if included in the circuit of Fig. 28, has two effects. First, the plate-cathode capacitance of the diode adds to C_t, thus increasing a little the rise time of the edges of reproduced pulses. Second, conduction current in the diode will alter the amplifier response whenever e_2 becomes negative. Negative values of e_2 are obtained during the overshoot in Fig. 31*B*, and the effect of the clamper on the overshoot is indicated by the dashed line in that figure.



(A) 0.5-μs PULSE (B) 500-μs PULSE

FIG. 31. Response of amplifier of Fig. 28 to rectangular pulses.

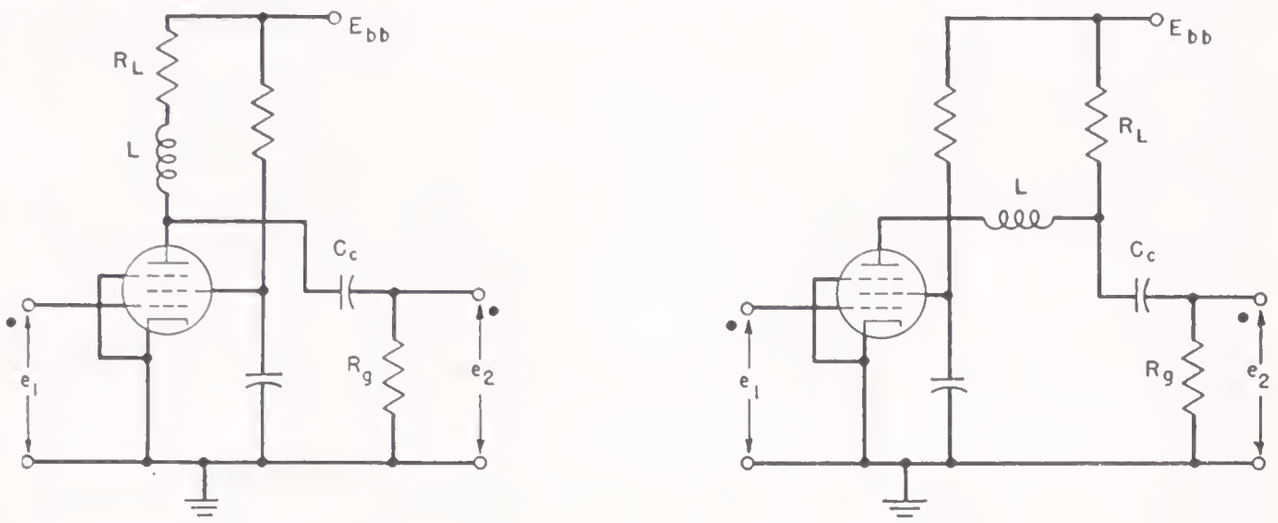
The time constant of pulse edges can be decreased, or the upper half-power frequency increased, only if either R_t or C_t is reduced. Capacitance C_t can never be less than the capacitance $C_o + C_i$ of the tubes employed, and reduction of R_t produces a corresponding loss of mid-band amplification. Extremely wide bands can therefore be obtained only through sacrifice of gain. An indication of the value of an amplifier is therefore its ability to provide both gain and bandwidth, and a convenient figure of merit is the *gain-bandwidth product*. For an R-C-coupled amplifier (without the cathode resistor of Fig. 28, which would be removed or by-passed if maximum gain-bandwidth were sought), the mid-band amplification is $g_m R_t$, and the bandwidth (essentially equal to the upper

half-power frequency) is $1/2\pi R_t C_t$. Thus

$$\text{Gain-bandwidth product} = g_m R_t \times \frac{1}{2\pi R_t C_t} = \frac{1}{2\pi}\frac{g_m}{C_t} \tag{38}$$

The value $g_m/(C_i + C_o)$ assumed by g_m/C_t if the wiring capacitance is ignored is considered a figure of merit for tubes to be used in wide-band amplifiers. The coefficient of g_m/C_t in Eq. 38 is a property of the circuit; the value $1/2\pi$ applies only to the basic R-C-coupled connection.

21. Compensated Video Amplifiers. The gain-bandwidth product obtainable with a specified tube type can be made greater than $(1/2\pi)g_m/C_t$ if an inductance is added to the circuit of the R-C-coupled amplifier. One such circuit, the *shunt-compensated* amplifier, is shown in Fig. 32*A*, and



(A) SHUNT COMPENSATED (B) SERIES COMPENSATED

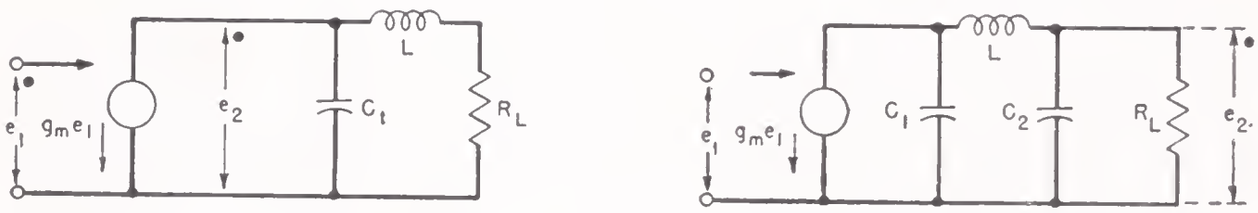
FIG. 32. Compensated video amplifiers.

another, a form of *series-compensated* amplifier, appears in Fig. 32*B*. The designations *shunt* and *series* indicate that in the first circuit the compensating inductance lies in a branch of the circuit shunted between the signal path and ground whereas in the second circuit the inductance is a series element in the path of the signal. More complicated circuits comprising a combination of series and shunt compensation or containing a number of inductances and capacitances in a configuration resembling an artificial transmission line may be employed, but in radar amplifiers only the series and shunt circuits are common.

The inductance in a shunt-compensated or series-compensated circuit has a small value, so that its reactance is appreciable only at high video frequencies. The frequency characteristics of the amplifiers in Fig. 32 are therefore like that of Fig. 30, except in the high-frequency range, and the pulse responses differ from those of Fig. 31 only in the neighborhood of the pulse edges. Equivalent circuits for determining high-frequency and pulse-edge response are drawn in Fig. 33. The circuits of this figure

are simplified in that coupling capacitors are considered to be short circuits and r_p and R_g are omitted because their resistances are large relative to the impedances they shunt. (Compare values of R_L, r_p, and R_g in Art. 20.) Capacitor C_t in Fig. 33*A* represents the sum of tube and wiring capacitances, as in Art. 20. Capacitor C_1 in Fig. 33*B* represents the component of C_t on the plate side of inductance L; C_2 represents the component on the grid side of the inductance.

The circuits of Fig. 33 are drawn in such a way as to illustrate the similarities between compensated amplifiers and artificial transmission lines, or low-pass filters. Figure 33*A* indicates that the shunt-compensated amplifier is equivalent to a current source $g_m e_1$ connected to the input terminals of an artificial line comprising a single L section and terminated by resistor R_L. The voltage developed across the input



(A) SHUNT COMPENSATED (B) SERIES COMPENSATED

FIG. 33. High-frequency (pulse edge) equivalent circuits of compensated amplifiers of Fig. 32.

impedance is e_2, the output voltage of the amplifier. Similarly, the series-compensated-amplifier equivalent circuit of Fig. 33*B* shows the current generator connected to an artificial line of one π section. The output voltage of the line is e_2, the output voltage of the amplifier. Were the artificial lines completely equivalent to real transmission lines, R_L could be made equal to the characteristic resistance R_c in each circuit, and the amplification would then be $g_m R_c$ for all frequencies. Actually, cutoff frequencies exist (see Art. 22, Chap. II) and limit the bandwidth, and even within the transmission band the amplification is not constant because of variations of the characteristic resistance.

The details of the performance of the shunt-compensated amplifier may be determined by calculating the amplification $g_m Z$, where Z is the complex-number impedance of L and R_L in parallel with C_t. The results may conveniently be expressed in terms of ω_h, the half-power angular frequency of the uncompensated circuit, and Q, the sharpness-of-resonance factor of the series R_L, L, C_t circuit. From Fig. 33*A* (above) and Art. 18, Chap. II,

$$\omega_h = \frac{1}{R_L C_t} \quad \text{and} \quad Q = \frac{\sqrt{L/C_t}}{R_L}$$

If the values of R_L and C_t are constant and the value of L is adjusted to

provide a desired type of response, then Q^2 may be considered a measure of the magnitude of L. Thus

$$Q^2 = \frac{L/C_t}{R_L^2} = \frac{L/R_L C_t}{R_L} = \frac{\omega_h L}{R_L} \tag{39}$$

that is, Q^2 is the ratio of the reactance of L (at the half-power frequency) to the load resistance.

The effects of the inductance on the frequency characteristic and on the pulse response are illustrated in Fig. 34. The $Q = 0$ curve of the amplitude-frequency plots in Fig. 34*A* corresponds to $L = 0$ and is like the

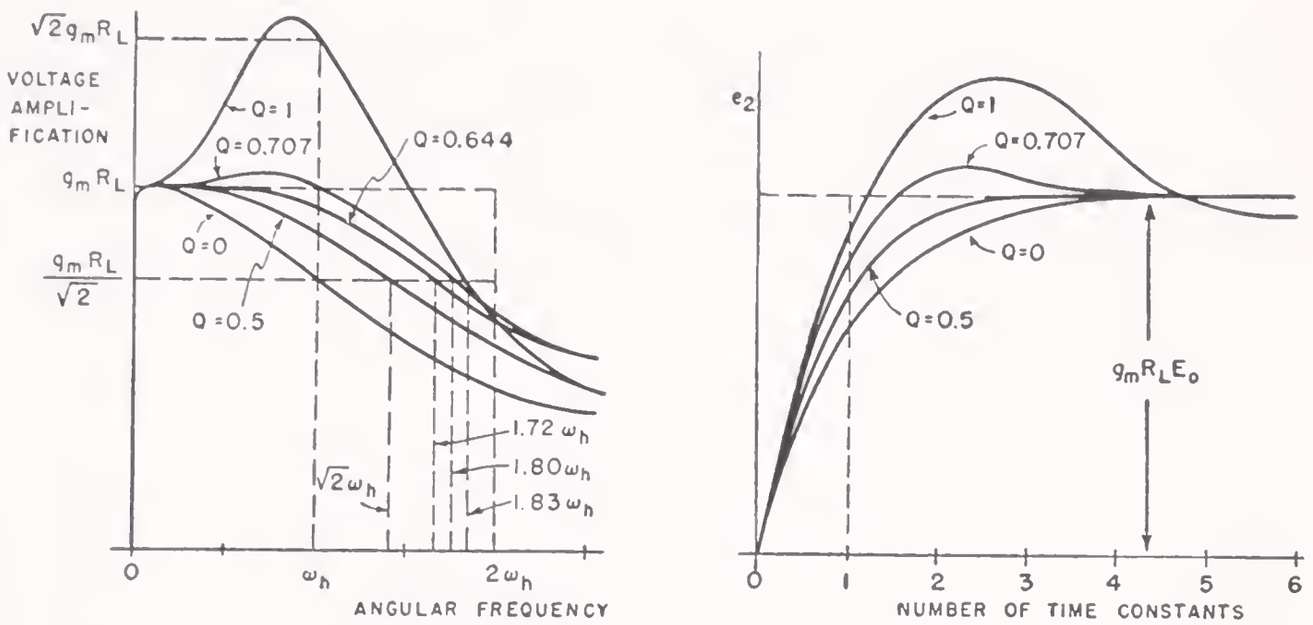


FIG. 34. Response of shunt-compensated amplifier.

curve of Fig. 30. Introducing L causes an inductive-susceptance component of the admittance of the R_L, L branch of the circuit in Fig. 33*A*. This inductive susceptance partly compensates the capacitive susceptance of C_t and increases the high-frequency response of the circuit, as indicated in Fig. 34*A*. For high values of L (corresponding to high values of Q), the resonance of L and C_t causes a peak in the response curve, as illustrated by the $Q = 1$ curve. The value $\sqrt{\sqrt{2} - 1} = 0.644$ for Q can be shown to provide transitional response—the greatest extension of bandwidth without a peak in the response curve.

For the curves of Fig. 34*B*, the grid voltage e_1 is assumed to be a negative pulse of amplitude E_o, and the $Q = 0$ curve corresponds to the leading edge of the e_2 pulse in Fig. 31*A*. The initial slope of all the curves in Fig. 34*B* is the same, this slope being determined by the rate at which current $g_m e_1$ can charge capacitor C_t (see Fig. 33*A*). For the $Q = 0$ curve, the rising voltage e_2 causes some of the source current to flow through R_L,

thus reducing the rate at which C_t charges and producing the exponential transient. With L present, the growth of current in R_L is delayed, and essentially the initial rate of rise of e_2 is maintained for a longer time. Large values of L are undesirable because of the oscillations that follow pulse edges. The transition between nonoscillatory and oscillatory response occurs for $Q = \frac{1}{2}$, corresponding to critical damping.

The value of Q employed in a shunt-compensated stage of a radar receiver depends upon the design of the other stages—the over-all frequency characteristic of the receiver should not be peaked nor have too abrupt a cutoff characteristic. As may be judged from Fig. 34, Q values of $\frac{1}{2}$ to 1 are typical. If, for example, a Q of $\frac{1}{2}$ is employed, the compensating inductance raises the half-power angular frequency from ω_h to $\sqrt{2}\omega_h$, and the gain-bandwidth product is increased from $(1/2\pi)g_m/C_t$ to $(1/\sqrt{2}\pi)g_m/C_t$. A somewhat greater increase of gain-bandwidth product is possible with the series-compensated circuit because the shunt capacitance is divided into two parts, C_1 and C_2.

22. Single-tuned Intermediate-frequency Amplifier. Video pulses to actuate the circuits of Arts. 19, 20, and 21 are derived by detection of intermediate-frequency pulses, which in turn are supplied by a high-gain i-f amplifier. The i-f circuits must respond to a band of frequencies centered about the intermediate carrier frequency—30 Mcps, for example—because the important Fourier components of the i-f pulses lie within such a band. The mid-band amplification required is of the order of a million, and the required bandwidth is 2 to 10 Mcps, depending upon the pulse duration and the purpose of the radar system (see Art. 18). To provide so much gain and bandwidth, i-f amplifiers must usually comprise at least six carefully designed stages.

A commonly used i-f circuit is the *single-tuned amplifier* illustrated by the circuit diagram of Fig. 35. Note that each stage includes only one tuning adjustment—inductance L is varied until resonance between it and the total shunt capacitance of the stage occurs at the desired intermediate frequency. The adjustment of L is usually accomplished by moving a core within the turns of the inductance. The core may be made of powdered magnetic material; if so, the inductance of the coil is increased when the core is inserted. More frequently, the core is a nonmagnetic metallic conductor such as a brass screw. Eddy currents in the metal surface prevent magnetic flux from entering the metal. Thus the magnetic field of the coil is restricted to the region outside the core, and the inductance of the coil is decreased by insertion of the core. The adjustable inductance is used for tuning in preference to an added adjustable capacitor because, as is to be shown, obtaining maximum gain and bandwidth requires maintaining the total capacitance at the minimum possible value.

The frequency characteristic and pulse response of each stage of the amplifier of Fig. 35 may be discussed in terms of equivalent circuits. An equivalent circuit for linear operation (Fig. 36*A*) is obtained directly from any one stage of Fig. 35 if by-pass capacitors C and coupling capacitor C_c are considered to have negligible reactance at the intermediate frequency. In Fig. 36*A*, $g_m E_1$, r_p, and output capacitance C_o represent

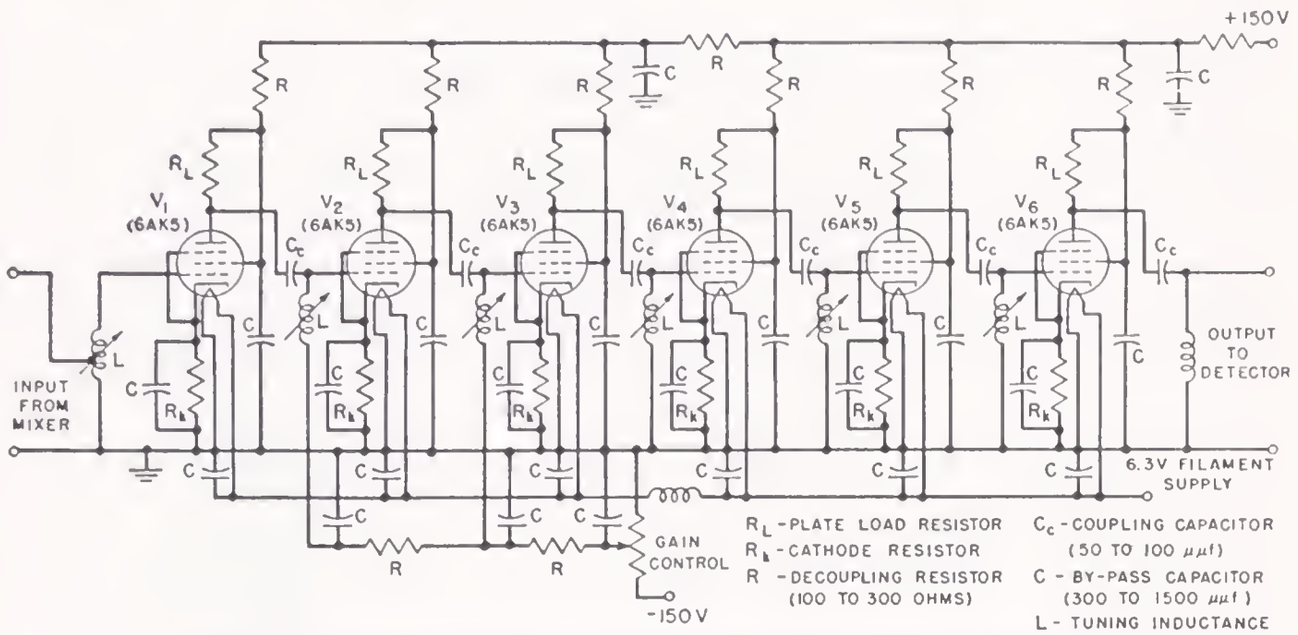


FIG. 35. Typical single-tuned intermediate-frequency amplifier.

the plate circuit of one of the Type 6AK5 tubes, and C_i and R_g (see Art. 9) represent the input impedance of the next stage. Because inductance L is not loss-free, resistor R_d is included in the equivalent circuit to represent coil dissipation. Capacitor C_w represents the sum of wiring capacitances and the interturn capacitance of the coil. Combination of the parallel resistors r_p, R_L, R_d, and R_g into R_t and of the parallel capacitors C_o, C_w, and C_i into C_t yields the simplified equivalent circuit of Fig. 36*B*.

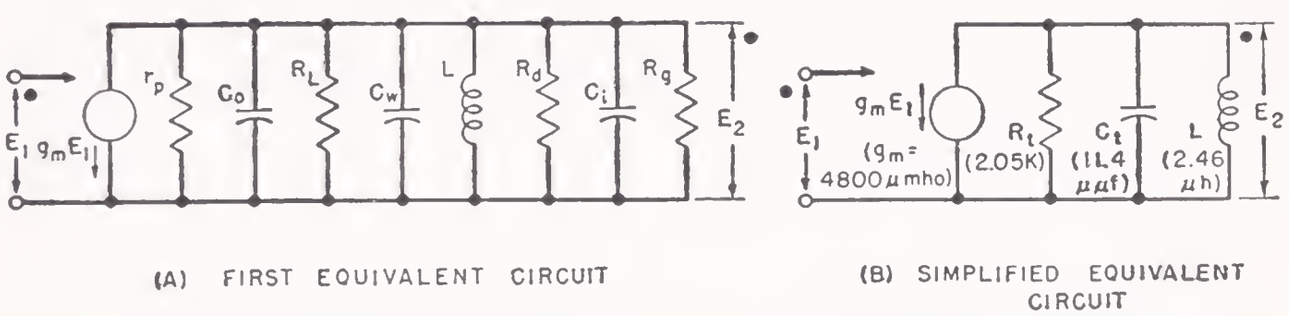


FIG. 36. Equivalent circuits for one stage of amplifier of Fig. 35.

Assume that, for the stage analyzed, the plate load resistor R_L in Fig. 35 is 2.3 K, the cathode resistor R_k is 200 ohms, and the decoupling resistors R are adjusted to provide a plate-and-screen-grid supply voltage of 120 volts. Reference to the characteristics of the 6AK5 shows the quiescent point to occur at a grid-bias voltage of approximately -2 volts. At this point the transconductance g_m is 4,800 μmhos, and r_p is approxi-

mately 500 K. For an intermediate frequency of 30 Mcps, the values $R_g = 30$ K and $R_d = 55$ K may be assumed. Thus R_t in Fig. 36*B* is the parallel combination of 500, 2.3, 55, and 30 K, or $R_t = 2.05$ K as indicated in the figure. The output and input capacitances of the 6AK5 are 2.1 and 4.3 $\mu\mu$f, respectively, and therefore, if C_w is assumed to be 5 $\mu\mu$f, C_t in Fig. 33*B* is 11.4 $\mu\mu$f. Because the inductance must resonate with C_t at 30 Mcps, L must be adjusted so that $1/\sqrt{LC_t}$ equals $2\pi \times 30$ Mcps; that is, L must equal 2.46 μh.

The output voltage E_2 in Fig. 36*B* is the product of the current from the source $g_m E_1$ by the impedance of R_t, C_t, and L in parallel, and the stage amplification is therefore g_m times this impedance. The nature of the

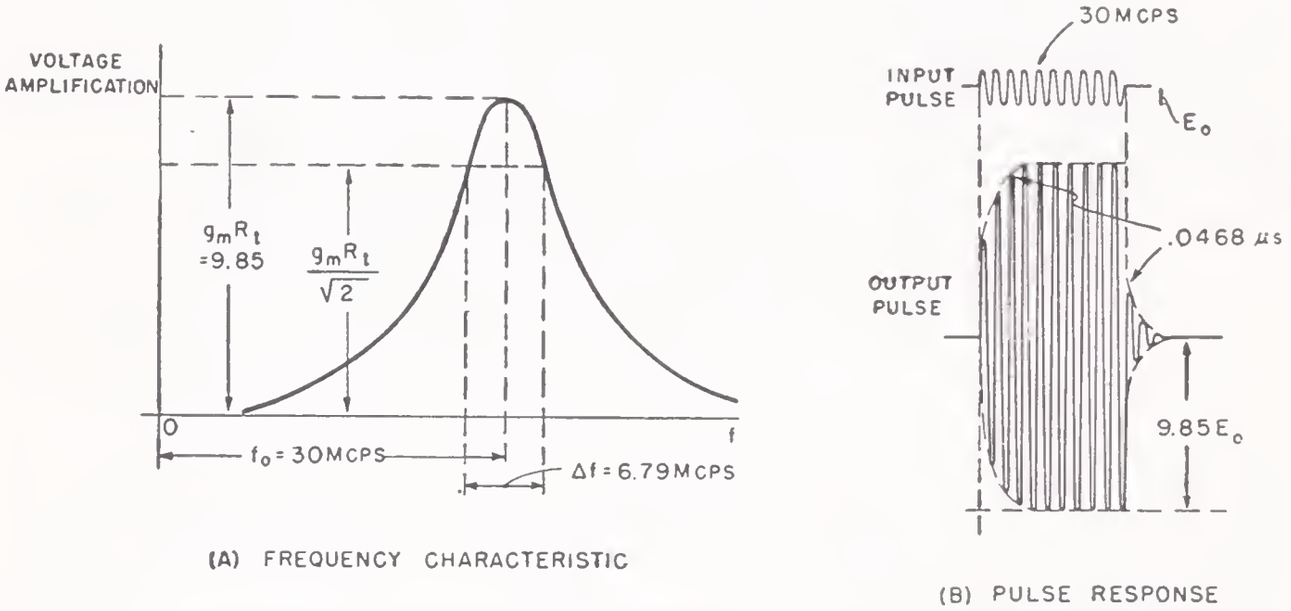


FIG. 37. Response of one stage of amplifier of Fig. 35.

impedance variation is shown by Fig. 7*C*, p. 352, and the frequency characteristic of the amplifier differs from the curve of this figure only by the scale factor g_m, as indicated in Fig. 37*A*. The mid-band amplification $g_m R_t$ is 4.8×2.05, or 9.85, and occurs, of course, at 30 Mcps. The bandwidth between half-power frequencies (see p. 355) is

$$\Delta f = \frac{1}{2\pi R_t C_t} = \frac{1}{2\pi \times 2050 \times 11.4 \times 10^{-6}} = 6.79 \text{ Mcps} \qquad (40)$$

The gain-bandwidth product $g_m R_t \times 1/2\pi R_t C_t$ is $(1/2\pi) g_m/C_t$ for the single-tuned i-f amplifier, this value being the same as for the *R-C*-coupled video amplifier. With respect to obtaining gain and bandwidth, therefore, the same quantity serves as a figure of merit for either i-f or video amplifiers.

The single-stage pulse response corresponding to the frequency characteristic of Fig. 37*A* may be determined from the principles explained in Art. 8 and has the form indicated in Fig. 37*B*. The video circuit used in

determining the envelope response of the i-f amplifier comprises the parallel combination of R_t and $2C_t$ connected to a current source equal to g_m times the grid-voltage envelope. Thus the exponential curves forming the edges of the output-pulse envelope in Fig. 37*B* have time constants

$$2R_tC_t = 2 \times 2{,}050 \times 11.4 \times 10^{-6} = 0.0468 \ \mu s$$

and the amplitude of the output pulse is $g_mR_t = 9.85$ times that of the input, provided the pulse duration is at least several tenths of a microsecond.

When several stages are cascaded, the over-all mid-band amplification is the product of the mid-band amplifications of the stages. Thus, in Fig. 35, if the value 9.85 applies to each of the six stages, the over-all amplification is $9.85^6 = 913{,}000$. Cascading stages to secure high amplification has the undesirable effect of reducing bandwidth. This effect results from the gradual decrease of amplification for each stage as the frequency departs from the mid-band value. At the half-power frequency of the over-all curve, each stage can be responsible for only a small reduction of amplification, because the product of all the reduction factors must be $1/\sqrt{2}$. Each of six identical stages, for example, can reduce the amplification only by the sixth root of $1/\sqrt{2}$. This root is 0.945, and this nearly unity value occurs close to the peak of the curve of Fig. 37*A*. Thus the over-all bandwidth between half-power points for a circuit such as Fig. 35 is much less than the corresponding bandwidth of the individual stages.

A measure of the amount of bandwidth reduction caused by cascading that is accurate if the per-unit bandwidth of individual stages is reasonably small may be had from Eq. 19, p. 356. The amplification of a single stage is g_m times the resonant-circuit impedance specified by Eq. 19, and thus $|A_1|^2$, the square of the magnitude of this amplification, is

$$|A_1|^2 = (g_mR_t)^2 \frac{1}{1 + [2(\omega - \omega_0)/\Delta\omega_1]^2}$$

where $\Delta\omega_1$ corresponds to $1/RC$ in Eq. 19 and is the angular-frequency bandwidth between half-power points for a single stage. If n identical stages are cascaded, the square of the magnitude of the over-all amplification $|A_n|^2$ is

$$|A_n|^2 = \frac{(g_mR_t)^{2n}}{\{1 + [2(\omega - \omega_0)/\Delta\omega_1]^2\}^n} \tag{41}$$

Let $\Delta\omega_n$ be the angular-frequency bandwidth between half-power points of the over-all frequency characteristic. Then at these half-power points, $\omega - \omega_0$ is $\pm\Delta\omega_n/2$ and $|A_n|^2$ has one-half of its mid-band value.

The denominator in Eq. 41 therefore equals 2, or

$$\left[1 + \left(\frac{\Delta\omega_n}{\Delta\omega_1}\right)^2\right]^n = 2$$

and

$$\frac{\Delta\omega_n}{\Delta\omega_1} = \frac{\Delta f_n}{\Delta f_1} = \sqrt{2^{1/n} - 1} \tag{42}$$

where Δf_n is the over-all bandwidth and Δf_1 the one-stage bandwidth. Values of the bandwidth-reduction factor are listed in Table 3. For the six-stage amplifier of Fig. 35, the reduction factor is 0.35, if the stages are identical, and the over-all bandwidth is 0.35 × 6.79, or 2.37 Mcps.

TABLE 3. BANDWIDTH-REDUCTION FACTORS FOR SINGLE-TUNED AMPLIFIERS

n	$\Delta f_n/\Delta f_1$	n	$\Delta f_n/\Delta f_1$
1	1.000	6	0.350
2	0.644	7	0.323
3	0.509	8	0.301
4	0.435	9	0.283
5	0.386	10	0.268

The reduction of bandwidth caused by cascading places an important limit upon the combinations of gain and bandwidth that can be attained with a single-tuned amplifier comprising identical stages. To increase the gain without loss of bandwidth, for example, stages must be added, and the bandwidth of each stage must therefore be increased. Consequently, the gain per stage decreases, and each additional stage provides a progressively smaller increment of gain. Eventually a limit is reached above which the gain is decreased by further addition of stages. The same sort of limit is reached if bandwidth is to be increased at constant over-all gain, and consequently the single-tuned circuit is useful only if the gain and bandwidth requirements are moderate. For example, in the example of Fig. 35, an amplification of nearly a million and a bandwidth of 2.37 Mcps are provided by a practical amplifier comprising six stages. The same gain and greater bandwidth could be obtained by utilizing more stages of the same kind, but 4.6 Mcps (corresponding to 10 stages) is about the practical limit, and 6.5 Mcps (corresponding to 27 stages) is the maximum possible bandwidth under the conditions specified.

Because of the high gain, high signal frequencies, and large bandwidth of radar i-f circuits, unintentional feedback, even in minute amounts, cannot be tolerated. The return to the input of as little as one-millionth of the output voltage may cause the circuit to operate as an oscillator, thus

destroying its usefulness as an amplifier. Feedback too small to cause oscillations may change the gain and frequency characteristic markedly.

To prevent unwanted feedback, extraordinary care in the design and construction of the amplifier is required. Decoupling elements, exemplified by resistors R and capacitors C in Fig. 35, are placed in plate and filament power-supply leads to prevent transmission of return signals along these leads. The entire amplifier is mounted in a separate, fully enclosed metal shield to prevent coupling by stray capacitances, and tubes of successive stages are spaced along a straight line to provide maximum distance between input and output circuits. The tuning

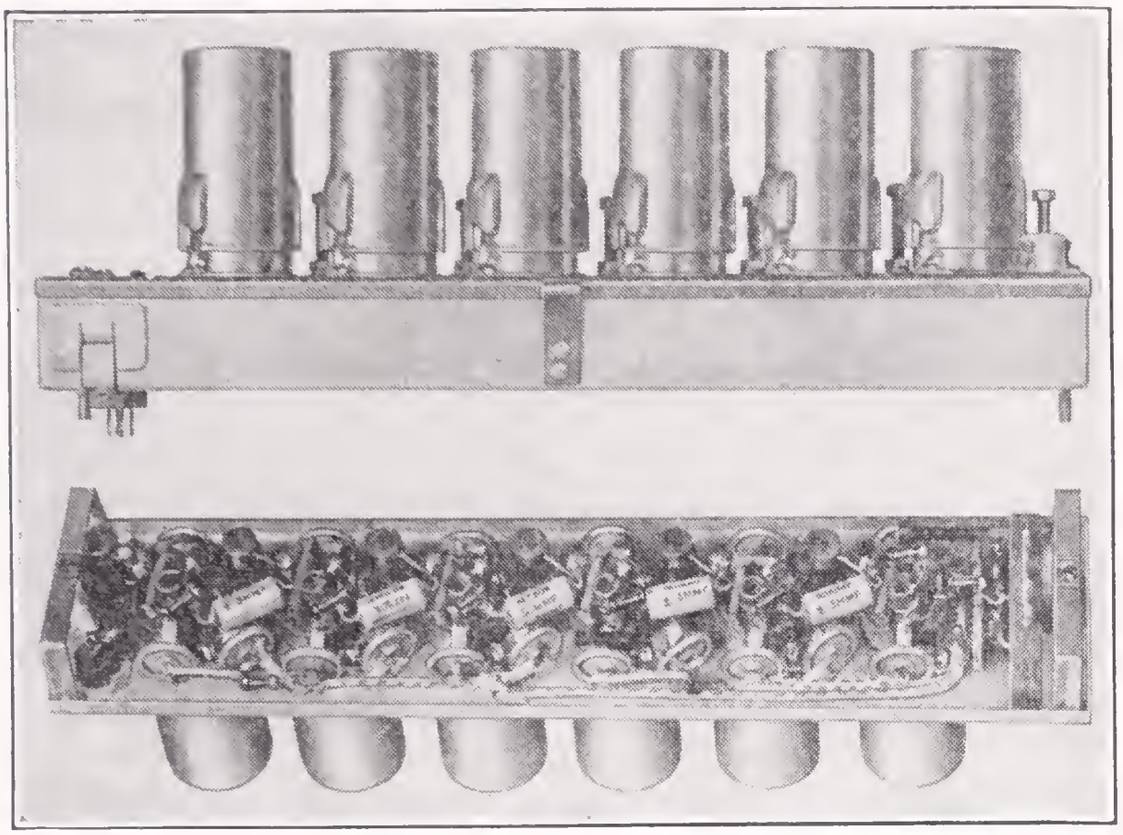


FIG. 38. Typical i-f amplifier. (*Courtesy of Bell Telephone Laboratories.*)

inductances are located and oriented in such a way as to minimize magnetic-field coupling. Output currents of each stage are returned directly to the cathode of that stage to prevent appreciable i-f currents in the metal chassis. Such currents can produce voltage drops between points of the chassis and thus provide unwanted components of grid voltage. Figure 38 illustrates these general features of construction. When all other causes of feedback are eliminated, there remains the possibility of the rectangular tube that encloses the circuit wiring operating as a waveguide to return signals from the output to the input. Such a waveguide operates at a frequency far below cutoff and attenuates greatly any signals that traverse it (see Art. 12, Chap. VIII). By spacing adequately the successive stages of the amplifier, the waveguide attenuation is made sufficient to prevent appreciable feedback.

Intermediate-frequency amplifiers, as well as video amplifiers, are subject to blocking because of the charges built up on coupling capacitors by large signals. To avoid such effects, the tuning inductances L in Fig. 35 are placed on the grid rather than the plate side of the coupling capacitors. With this connection, a change of coupling capacitor voltage affects mainly the plate voltage preceding the capacitor. The use of a resistor instead of an inductor in the plate circuit has the additional effect in the circuit of Fig. 35 of providing a quiescent plate voltage lower than the screen-grid voltage. In this way a very high plate current may be obtained without exceeding the allowable plate dissipation. The transconductance g_m (which depends chiefly on plate current) is thus increased, and a corresponding improvement in gain-bandwidth is obtained.

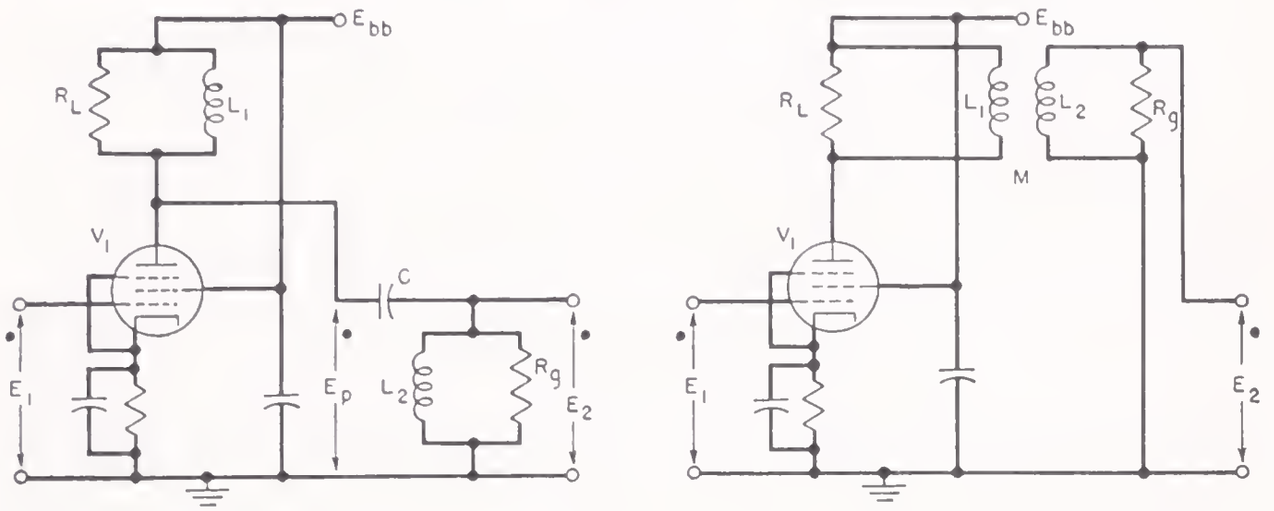
Additional possibilities for blocking exist because of the screen-grid and cathode by-pass capacitors. During the application of large signal voltages, the screen-grid and cathode currents are increased, resulting in a decrease of screen-grid voltage and increase of control-grid-bias voltage. The gain is thereby reduced if the by-pass capacitors maintain the abnormal voltages after the signal is removed. To minimize the duration of the reduced-gain interval, the by-pass capacitors are chosen as small as possible consistent with maintaining negligible i-f reactance.

Because echo pulses from large, near-by objects and from small, distant objects differ tremendously in strength, a gain control in the i-f circuit is essential. In Fig. 35, the gain control adjusts the grid-bias voltage for tubes V_2 and V_3. Changing the grid-bias voltage in the negative direction reduces the plate current and the transconductance at the operating point; in consequence, the small-signal gain is decreased. The gain control is always applied to stages near the amplifier input in order to prevent saturation of stages ahead of the control point by the largest possible signals. If the i-f amplifier is part of a receiver employing a crystal mixer and no r-f amplifiers, gain control usually begins with the second stage. Signals in the first stage do not reach saturation amplitude, and the design of the first stage for minimum noise figure (see Arts. 17 and 25) is simplified if provision for gain control is not required. For similar reasons, in receivers having an r-f amplifier, gain control usually begins with the second r-f stage.

If the gain-control voltage were applied only to V_2 in Fig. 35, the range of gain control would be inadequate, because stray grid-plate coupling (for example, that caused by grid-plate capacitance) causes the stage to transmit a small part of the grid signal even when plate current in tube V_2 is cutoff. By applying the gain-control voltage to two, or sometimes three, consecutive stages, sufficient gain reduction may be obtained to provide normal-amplitude output with the largest possible input signals. Use of large bias voltages to control gain in the presence of large

input signals may result in a highly nonlinear amplification characteristic, the controlled tubes carrying current only on the peaks of the large signals and being entirely cutoff for small signals. Such an effect would cause intolerable distortion in an amplitude-modulation communications receiver, but it is not necessarily a disadvantage in radar receivers. It can be avoided by applying the gain-control voltage to more stages, so that bias voltages of smaller amplitude can be employed in each stage.

In some radar receivers the screen-grid voltage rather than the grid-bias voltage is adjusted to achieve gain control. The grid-bias form of control is usually preferable because it protects the controlled stage from saturation by large signals. Screen-grid voltage control is worse than no protection for the controlled stage—reducing the screen-grid voltage permits smaller amplitude signals to saturate the stage.



(A) CAPACITIVE COUPLING (B) INDUCTIVE COUPLING

FIG. 39. Double-tuned amplifier stages.

23. Double-tuned Coupling Circuits. To avoid the gain and bandwidth limitations of single-tuned i-f amplifiers, coupling circuits of other forms are sometimes employed, especially where very wide-band amplification is required. One type of coupling network illustrated by Figs. 39*A* and *B* utilizes two resonant circuits per stage and is thus *double tuned*. In each of the figures, L_1 and L_2 together with tube and wiring capacitances form the two resonant circuits. In Fig. 39*A*, the two circuits are coupled by a small capacitor C; in Fig. 39*B*, a loosely coupled transformer provides not only inductances L_1 and L_2 but also mutual-inductance coupling.

The double-tuned circuit, like the single-tuned one, provides response to a band of frequencies surrounding a carrier frequency. Parameters of the circuit control the mid-band frequency, the mid-band amplification, and the bandwidth; and in contrast to the single-tuned circuit, the basic shape of the response characteristic can be altered by circuit-parameter adjustments. To demonstrate the relations involved, the complex-

number amplification is determined from a linear equivalent circuit and placed in a form convenient for interpretation.

Figure 40*A* is an equivalent of Fig. 39*A*. The parallel combination of R_L with the plate resistance of V_1 and the loss resistance of L_1 is denoted by R_1. Similarly, the parallel combination of grid-circuit resistances is R_2, and the total plate-circuit and grid-circuit capacitances are C_1 and C_2. The admittances Y_1, Y_2, and Y are defined in the figure. The cur-

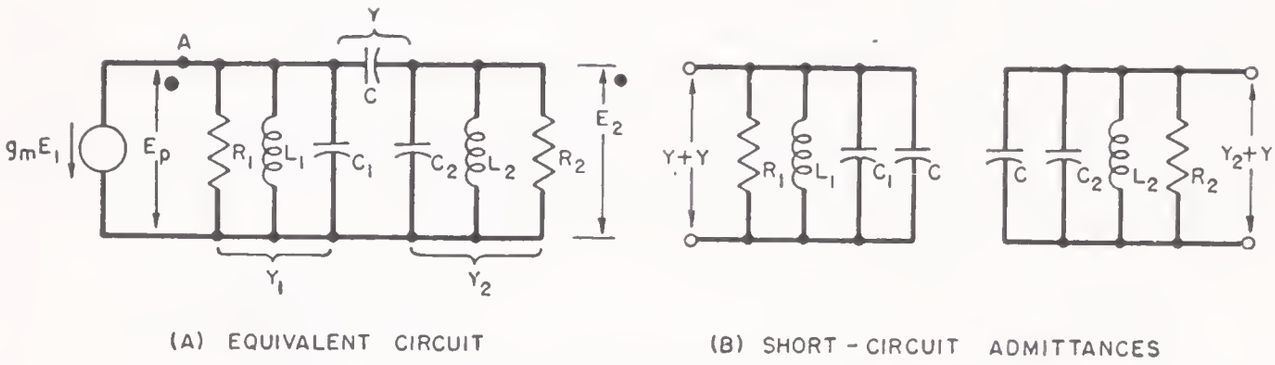


FIG. 40. Linear equivalent circuit for Fig. 39*A*.

rent $-g_mE_1$ entering the coupling network at point A is the sum of currents in Y_1 and Y, and the current in Y equals that in Y_2; that is,

$$\begin{aligned} -g_mE_1 &= Y_1E_p + Y(E_p - E_2) \\ Y(E_p - E_2) &= Y_2E_2 \end{aligned} \tag{43}$$

Solving the second equation for E_p, substituting the value obtained into the first equation, and solving the result for E_2/E_1 yields an expression for the amplification

$$A = \frac{E_2}{E_1} = -g_m \frac{Y}{(Y_1 + Y)(Y_2 + Y) - Y^2} \tag{44}$$

The assumption is now made that the per-unit bandwidth is small. Thus Y is approximately constant at the value $j\omega_0 C$, ω_0 being the mid-band angular frequency. Also, the short-circuit admittances $Y_1 + Y$ and $Y_2 + Y$, which pertain to parallel R-L-C circuits adjusted to resonate at ω_0, may be expressed in terms of the small-bandwidth approximation of Art. 6 (see p. 356) as

$$\begin{aligned} Y_1 + Y &= \frac{1}{R_1}\left[1 + j2\frac{Q_1}{\omega_0}(\omega - \omega_0)\right] \\ Y_2 + Y &= \frac{1}{R_2}\left[1 + j2\frac{Q_2}{\omega_0}(\omega - \omega_0)\right] \end{aligned} \tag{45}$$

where Q_1 is $\omega_0(C_1 + C)R_1$ and Q_2 is $\omega_0(C_2 + C)R_2$. These admittance values are substituted into Eq. 44, and the following new symbols are introduced:

$$k = \frac{C}{\sqrt{(C_1 + C)(C_2 + C)}} \tag{46}$$

$$\Delta\omega = \omega_0 \sqrt{k^2 + \frac{1}{Q_1 Q_2}} \tag{47}$$

$$B = \Delta\omega \sqrt{(C_1 + C)(C_2 + C)} \tag{48}$$

Quantity k is the *coefficient of coupling* for the capacitor network comprising C_1 and C_2 joined by C and can range from zero to 1. Quantity $\Delta\omega$ is a measure of bandwidth to be interpreted shortly. Thus $\sqrt{k^2 + 1/Q_1Q_2}$ is per-unit bandwidth, and for small per-unit bandwidth k must not range far from zero. Quantity B is the susceptance of the geometric mean of the short-circuit capacitances at the angular frequency $\Delta\omega$. In terms of these symbols, Eq. 44 becomes, after some manipulation,

$$A = \frac{g_m}{jB} \frac{k}{\sqrt{k^2 + 1/Q_1Q_2}} \times \frac{1}{1 - [2(\omega - \omega_0)/\Delta\omega]^2 + j[(1/Q_1 + 1/Q_2)/\sqrt{k^2 + 1/Q_1Q_2}][2(\omega - \omega_0)\Delta\omega]} \tag{49}$$

The first two fractions in Eq. 49 are independent of frequency and represent mid-band amplification. The third fraction (which equals unity at mid band, $\omega = \omega_0$) indicates the effect of frequency on amplification. Note that change of ω from $\omega_0 - \frac{1}{2}\Delta\omega$ through ω_0 to $\omega_0 + \frac{1}{2}\Delta\omega$ changes the phase of the third fraction from $+90$ deg through zero to -90 deg. Thus $\Delta\omega$ is the angular-frequency bandwidth between points of 90-deg phase shift.

The first two fractions determine only the vertical scale of the response curve. The bandwidth $\Delta\omega$ is a scale factor for the width of the response curve, corresponding to $\omega_0/2Q$ for the single-tuned resonance curve. The quantity $(1/Q_1 + 1/Q_2)/\sqrt{k^2 + 1/Q_1Q_2}$, found in the third fraction, determines the shape of the response curve. The effect of this quantity on the amplitude response $|A|$ may be seen from a study of the denominator of the third fraction, which denominator is proportional to $1/A$ and is

$$D = 1 - \left(2\frac{\omega - \omega_0}{\Delta\omega}\right)^2 + j\frac{1/Q_1 + 1/Q_2}{\sqrt{k^2 + 1/Q_1Q_2}}\left(2\frac{\omega - \omega_0}{\Delta\omega}\right)$$

and thus

$$|D| = \sqrt{1 + \eta\left(2\frac{\omega - \omega_0}{\Delta\omega}\right)^2 + \left(2\frac{\omega - \omega_0}{\Delta\omega}\right)^4} \tag{50}$$

where

$$\eta = \frac{1/Q_1^2 + 1/Q_2^2 - 2k^2}{k^2 + 1/Q_1Q_2} \tag{51}$$

is the *shape factor* of the response. If η is positive, any value of ω different from ω_0 makes $|D|$ greater than 1. If $\eta = 0$, the square term disappears from Eq. 50 and $|D|$ is nearly unity for all values of ω near ω_0 but increases abruptly when ω approaches $\omega_0 \pm \Delta\omega/2$. If η is negative, the square term becomes negative and $|D|$ is less than 1 for small variations of ω from ω_0 but greater than 1 (because of the fourth-power term) for large variations. The effect of these variations of $|D|$ on the amplitude-response curve is indicated by the curves of Fig. 41. (Scale relations among the curves for different values of η are as indicated only for the conditions specified in the legend of the figure. The curve shapes, however, depend only upon the value of η.)

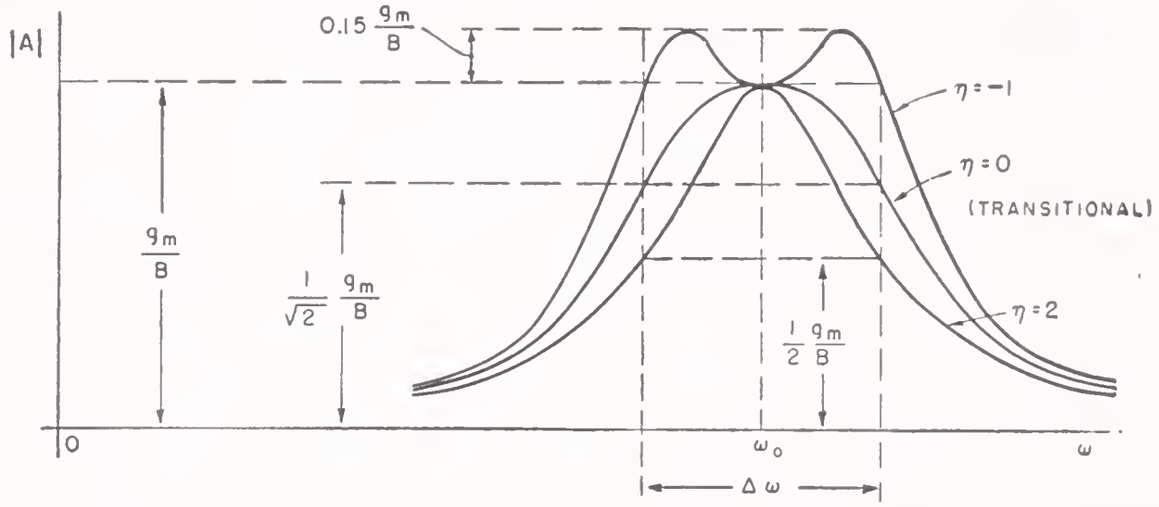


FIG. 41. Amplitude-response curves for double-tuned coupling circuit. (Drawn for fixed $k = \Delta\omega/\omega_0$ and either $1/Q_1$ or $1/Q_2$ equal to zero.)

The condition $\eta = 0$ is called *transitional coupling*, because it separates the condition of double-peaked response (overcoupling) and narrow single-peaked response (undercoupling). Equation 51 indicates the condition for transitional coupling to be

$$k^2 = \frac{1}{2}\left(\frac{1}{Q_1^2} + \frac{1}{Q_2^2}\right) \tag{52}$$

From the curves of Fig. 41, it may be seen that $\Delta\omega$ is the bandwidth between half-power points *for transitional coupling only*. For constant $\Delta\omega$, the half-power bandwidth increases as the coupling is increased.

Comparison of Fig. 34*A*, p. 411, with Fig. 41 suggests that the double-tuned i-f amplifier is related to the compensated video amplifier in the same way that the single-tuned amplifier is related to the basic *R-C*-coupled circuit. According to the general principle of Art. 8, therefore, the effect of double-tuned circuits on pulse envelopes can be inferred from knowledge of the pulse response of compensated video amplifiers. Overcoupling corresponds to overcompensation and produces oscillations of the envelope curve in the vicinity of pulse edges. Thus, if all stages of an i-f amplifier are double tuned, coupling greater than transitional should

be avoided. An amplifier comprising only double-tuned stages is difficult to adjust, and therefore double-tuned stages are ordinarily employed in combination with a greater number of single-tuned stages. In such an amplifier, the double-tuned stages may be overcoupled; the combination of double-peaked and single-peaked response curves yields a flat over-all characteristic, and the high band-edge amplification of the double-tuned stages effects an increase of the gain-bandwidth product.

To provide maximum amplification when the bandwidth and response-curve shape are prescribed for a double-tuned circuit, the first two fractions in Eq. 49 should be maximized. The first fraction is essentially fixed by the possible values of g_m, C_1, and C_2, and the denominator of the second fraction is determined by the required per-unit bandwidth. The factor k remains to be maximized. Because $k^2 + 1/Q_1Q_2$ is fixed, *either* $1/Q_1$ *or* $1/Q_2$ should be made zero and the other Q quantity adjusted to provide the required value of η. This adjustment, assumed in the plots of Fig. 41, causes $k/\sqrt{k^2 + 1/Q_1Q_2}$ to be unity. For the maximum-gain adjustment and transitional coupling (where $\Delta\omega$ is the half-power bandwidth), therefore,

$$\text{Gain-bandwidth product} = \frac{g_m}{B}\frac{\Delta\omega}{2\pi} \approx \frac{1}{2\pi}\frac{g_m}{\sqrt{C_1C_2}}. \tag{53}$$

capacitor C being ignored relative to C_1 and C_2 because k is small. For the single-tuned amplifier, $C_t = C_1 + C_2$ or, if C_1 and C_2 are nearly equal, $C_t \approx 2\sqrt{C_1C_2}$. Thus the gain-bandwidth product of a single stage can be approximately doubled by the use of double tuning with transitional coupling and maximum gain.

A double-tuned coupling network with appreciable damping in one circuit only, as required for maximum gain, is difficult to realize, because power loss in both plate and grid circuits is unavoidable. Furthermore, with very slight damping of one circuit, the network becomes unduly sensitive to small changes of tube capacitance. As an indication of the loss of gain caused by practical operating conditions, consider transitional coupling with $Q_1 = Q_2 = Q$. Equation 52 becomes $k = 1/Q$, and the second fraction of Eq. 49 becomes $1/\sqrt{2}$, so that the amplification is about 0.7 of the maximum and the gain-bandwidth product is $\sqrt{2}$ times that obtainable with single-tuned coupling.

A greater increase of gain-bandwidth product is obtained if double-tuned coupling is used in an amplifier comprising several stages because the amplification within the response band is more nearly constant, so that the bandwidth reduction caused by cascading is less. For example, if six transitionally coupled double-tuned stages each have the same bandwidth as six single-tuned stages, the over-all amplification of the double-tuned stages is greater than that of the single-tuned stages by a

factor of 2^3 to 2^6 (dependent upon the ratio of Q values for the double-tuned circuits). The bandwidth reduction factor for the six double stages is 0.593 instead of 0.350 (Table 2, p. 416), so that an additional bandwidth-improvement factor of 1.7 is obtained with double tuning. For overcoupled double-tuned stages used in conjunction with single-tuned stages, the gain-bandwidth product is appreciably greater than for transitional coupling. Moreover, narrow-band single-tuned stages providing high gain may be employed, the over-all characteristic being widened by the effect of the double-peaked response of the double-tuned stages.

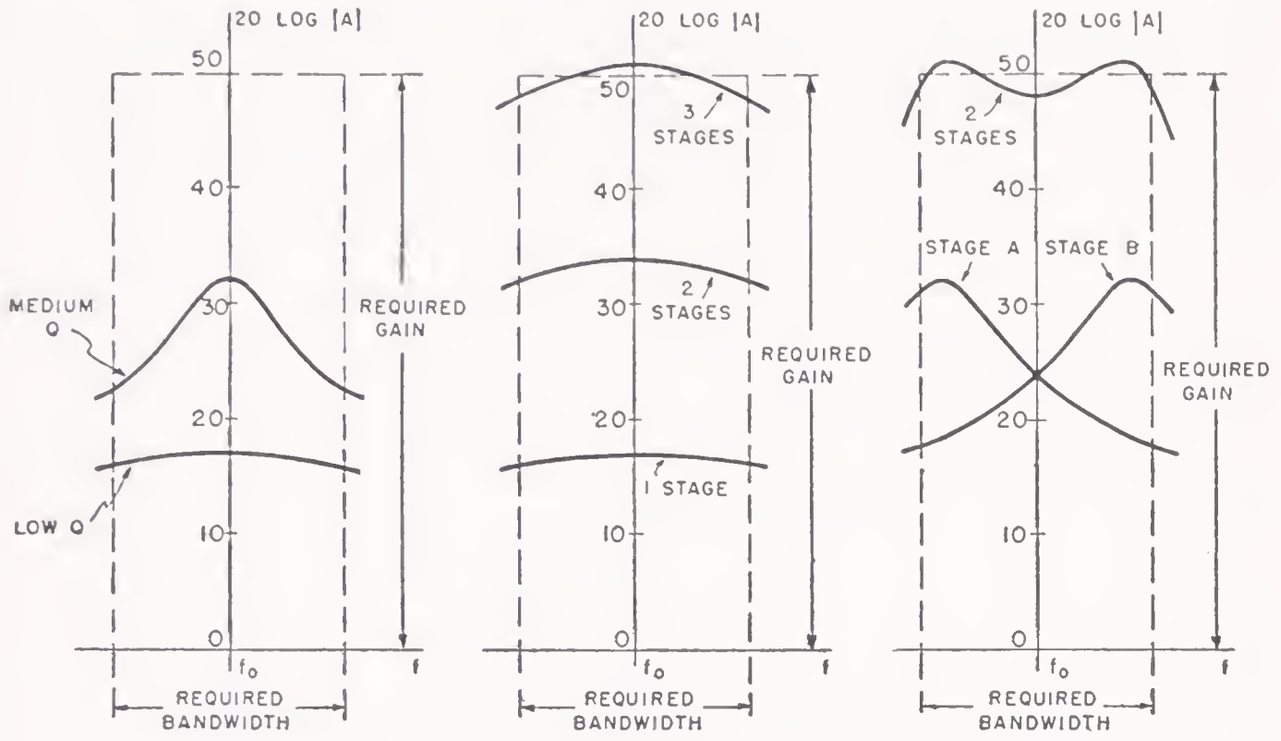
The above analysis, made for the capacitively coupled double-tuned circuit of Fig. 39*A*, applies also to the inductively coupled circuit of Fig. 39*B*. When the coupling is inductive, k in Eq. 49 is the usual coupling coefficient of a transformer, $M/\sqrt{L_1L_2}$ (see Eq. 32, p. 129, in Chap. II), and capacitor C is, of course, omitted from the expressions defining $\Delta\omega$ and B. Quantities Q_1, Q_2, and ω_0 pertain to the *short-circuit* admittances in which the inductances are $L_1(1 - k^2)$ and $L_2(1 - k^2)$ (see p. 129, in Chap. II). It is customary, nevertheless, to compute ω_0 and the Q values from L_1 and L_2, the results being negligibly changed because small per-unit bandwidth requires that k be much less than unity.

24. Stagger-tuned Amplifier. A second method of increasing the gain and bandwidth capabilities of i-f amplifiers does not require a change from the circuit configuration of Fig. 35. Instead, the improved performance is obtained by a change in tuning adjustments—the resonant frequencies for the various stages are distributed through the frequency band to be amplified. The product of the amplitude-response curves of all the stages is the over-all response curve; for this over-all curve to possess reasonably uniform and very high gain in a specified response band, the resonant frequencies and bandwidths of the individual stages must be selected in a special manner.

The reason that stagger tuning can improve gain-bandwidth performance may be seen in terms of Fig. 42. In this figure voltage gain is plotted in *decibels* ($20 \log |A|$) in order that over-all gain may be obtained by *adding* stage gains. Assume that an amplifier is required to have nearly uniform gain over a prescribed bandwidth and that the required values of gain and bandwidth are as indicated by the dashed lines. The area within the dashed-line rectangle may be considered a measure of the desired amplifier performance. A single stage employing a load resistor of the order of 10 K in shunt with the tuned circuit has the response characteristic indicated by the medium-Q curve in Fig. 42*A*. The area under this curve and within the required band is more than half the required area. Two such stages in cascade are therefore more than adequate to provide the required gain-bandwidth area, but the amplifica-

tion thus obtained would vary 20 db (a voltage ratio of 10:1) between the band center and the band edges.

To secure more nearly uniform response with aligned stages, the load resistors must be reduced to provide response curves such as the low-Q curve of Fig. 42*A*. Three such stages in cascade meet the requirements, as indicated in Fig. 42*B*. Such a three-stage amplifier is wasteful of gain-bandwidth area; in each stage the area between the two curves of Fig. 42*A* is discarded because of the small value of load resistance used. More effective use of the tube capabilities is secured if the higher value load resistance is employed and two stages of different resonant frequencies are cascaded. Figure 42*C* indicates the way in which *two*



(A) SINGLE-STAGE CURVES (B) ALIGNED STAGES (C) STAGGER-TUNED STAGES

FIG. 42. Response curves for cascaded single-tuned amplifiers.

stagger-tuned stages can be employed to meet the same overall response requirements as *three* low-Q aligned stages.

The general principle pertinent to this example is that optimum gain-bandwidth performance is obtained if all load resistors are chosen as large as operating conditions permit—practical limitations may be grid loading in the subsequent stage or the onset of feedback oscillations. Stagger-tuned groups comprising a sufficient number of stages to meet the bandwidth requirement are then formed, and enough of these groups are cascaded to fulfill the gain requirement. For narrow-band amplification, a single stage provides adequate bandwidth and completely aligned amplifiers are appropriate. For the widest bands employed in radar, however, stagger-tuned pairs or triplets are required for best utilization of tubes.

The possible shapes of the response curves for stagger-tuned pairs of small per-unit bandwidth are identical with those of double-tuned circuits. This fact may be demonstrated if the small-bandwidth approximation is employed in writing the amplifications of stages A and B in a stagger-tuned pair (as explained in Art. 22, the amplification is a scale factor times the impedance of a parallel R-L-C circuit):

$$A_A = \frac{A_{A0}}{1 + j2(\omega - \omega_A)/\Delta\omega_1}$$
$$A_B = \frac{A_{B0}}{1 + j2(\omega - \omega_B)/\Delta\omega_1} \tag{54}$$

where A_{A0} and A_{B0} are the mid-band amplifications, ω_A and ω_B are the staggered angular resonant frequencies, and $\Delta\omega_1$ is the half-power angular frequency bandwidth for each of the stages. The over-all amplification of the pair is $A = A_A A_B$. After this multiplication is performed, the following new symbols are introduced into the product expression:

$$\omega_0 = \frac{\omega_A + \omega_B}{2} \tag{55}$$

$$\omega_d = \omega_A - \omega_B \tag{56}$$

and

$$\Delta\omega = \sqrt{(\Delta\omega_1)^2 + \omega_d{}^2} \tag{57}$$

Combination of Eqs. 55 and 56 yields

$$\omega_A\omega_B = \omega_0{}^2 - \left(\frac{\omega_d}{2}\right)^2$$

and the final form obtained for the over-all amplification is

$$A = \frac{A_{A0}A_{B0}}{1 + (\omega_d/\Delta\omega_1)^2} \frac{1}{1 - [2(\omega - \omega_0)/\Delta\omega]^2 + j[2\Delta\omega_1/\Delta\omega][2(\omega - \omega_0)/\Delta\omega]} \tag{58}$$

The first fraction in Eq. 58 is the mid-band amplification, and the second fraction expresses the variation of A with frequency. A comparison of Eq. 58 with Eq. 49 (p. 421) shows the response-curve shapes for stagger-tuned pairs and double-tuned circuits to be the same. The quantity ω_0 in Eq. 58, as in Eq. 49, indicates the mid-band angular-frequency of the over-all curve, and $\Delta\omega$ indicates the angular-frequency bandwidth between points of 90-deg phase shift. The factor $2\Delta\omega_1/\Delta\omega$ in Eq. 58 determines the shape of the response curves. By computing the magnitude of the second-fraction denominator, as was done in obtaining Eq. 51, p. 421, the shape factor for stagger-tuned pairs can be found to be

$$\eta = 2\frac{1 - (\omega_d/\Delta\omega_1)^2}{1 + (\omega_d/\Delta\omega_1)^2} \tag{59}$$

Thus transitional response is obtained for $\omega_d/\Delta\omega_1 = 1$, or for a resonant-frequency spacing equal to the bandwidth. Wider spacing yields double-peaked response curves; closer spacing yields sharply peaked curves.

Somewhat lengthy calculations are required to determine the response of a group of three or more stages all tuned to different frequencies.[1] The features of the response characteristics are not altered. For wide spacing of resonant frequencies, the over-all curve contains as many peaks as there are stages in the group. Decreasing the spacing smooths the response curve until the nearly flat-topped transitional characteristic is obtained. With still closer spacing, the response curve tends toward that of aligned stages. When more than two stages are grouped, a critical choice of bandwidths and resonant-frequency intervals is required if the amplification within the response band is to be approximately uniform.

25. Intermediate-frequency Input Stages. The most critical stage of an i-f amplifier is the input stage. Upon the excellence of the design of this stage depends the noise figure of the receiver and the performance of the entire radar system with respect to detection of small objects at great range. Particularly are these statements true for receivers employing no r-f amplifiers because of the very low signal level at the input of the i-f amplifier. Not only must gain and bandwidth be considered in the design of the first i-f stage, but also, and most important, noise generation in this stage must be reduced as far as possible.

In the typical i-f amplifier of Fig. 35, p. 413, an input circuit of simple form is indicated. The mixer may be considered (see Art. 28) equivalent to a current source, a resistor, and a capacitor connected in parallel between ground and the tap on the inductance in Fig. 35. Noise and signal voltages are delivered to the i-f-amplifier terminals by the mixer, and additional noise is generated in the amplifier components, particularly tube V_1. As explained in Art. 17, the position of the tap on the input inductance controls the step-up ratio a between the mixer output circuit and the grid circuit of tube V_1. A particular adjustment of a yields at mid-band frequency a match between mixer resistance and grid-circuit resistance and supplies maximum signal voltage to the tube. Because of the noise generated in the grid circuit, a somewhat smaller value of a provides a minimum noise figure, and this smaller value is used.

If the bandwidth of the input circuit is much greater than the over-all bandwidth of the i-f amplifier, the input circuit operates under essentially mid-band conditions for all frequencies that can produce appreciable output, and the noise figure is constant at the mid-band value. If, on the other hand, the input bandwidth is restricted, signal components at fre-

[1] G. E. VALLEY, JR., and H. WALLMAN, *Vacuum Tube Amplifiers*, 166–200, 278–287 (McGraw-Hill Book Company, Inc., New York, 1948).

quencies near the band edge reach the input tube in reduced magnitude. Thus the importance of the plate-circuit noise is increased, and the noise figure is increased at the band edges. The integrated noise figure over the response band, which is the quantity of greatest interest, is in consequence somewhat increased by restriction of the input bandwidth.

In the circuit of Fig. 35, the input-circuit bandwidth depends upon the resistance obtained by combining the input resistance of tube V_1 in parallel with a^2 times the mixer resistance and upon the capacitance that is the parallel combination of the tube input capacitance and $1/a^2$ times the mixer output capacitance. If a reasonable attempt is made to minimize the capacitances, the tube and mixer resistances are ordinarily low enough to provide input-circuit bandwidths much greater than the over-all bandwidth of the i-f amplifier. If the shunt capacitances are very large, or if extremely wide response bands are required, the input-circuit bandwidth in Fig. 35 may be inadequate. A double-tuned input circuit can then be used to advantage because of its greater bandwidth. The coefficient of coupling is adjusted to provide the desired relation between the tube input resistance and the resistance it faces at the mid-band frequency. Note that a shunt resistor must not be added to the single-tuned circuit to increase the bandwidth in an effort to reduce the integrated noise figure. The resistor would increase the noise figure at all frequencies, both because it is a source of additional noise and because it reduces the signal voltage at the tube.

In Art. 16 it is pointed out that the equivalent noise resistance of a triode is much less than that of a pentode. It would seem, therefore, that an appreciable reduction of the noise figure might be had by the use of a triode rather than a pentode as the first stage in the amplifier of Fig. 35. The difficulty with using a triode is that at radar intermediate frequencies the feedback through the grid-plate capacitance is so great that oscillations occur whenever the load resistance is made large enough to provide useful voltage amplification.

This difficulty may be stated in a manner appropriate to noise-figure analysis: The noise figure of the first two stages is $F_1 + (F_2 - 1)/G_1$ (see Eq. 35, p. 394), where F_1 is the noise figure and G_1 the available-power gain of the first stage and F_2 is the noise figure of the second stage. For a triode i-f amplifier, G_1 may have a value of several hundred. Therefore if F_2 can be given a reasonably low value, $(F_2 - 1)/G_1$ can be made negligible and the over-all noise figure reduced essentially to F_1, the low noise figure of the input triode. Unfortunately, with the usual cathode-return form of amplifier as the second stage, reasonably low values of F_2 can be had only if the signal source for the second stage has an impedance somewhat lower than the input impedance of the second stage. In order to prevent oscillation of the triode first stage, the coupling network

between first and second stages must be adjusted so that the first stage is effectively a high-impedance source connected to a low impedance load. With this adjustment, high values result for F_2, and the quantity $(F_2 - 1)/G_1$ constitutes an appreciable and very undesirable increase of the noise figure.

This difficulty can be avoided if a second-stage amplifier can be provided for which a low value of F_2 is obtained when impedances are adjusted in the manner required to avoid oscillations. Such an amplifier is the *grid-return* or *grounded-grid* circuit employed in the second stage of Fig. 43*A*. (The significance of the designations *cathode return*, *grid return*, and *plate return* are discussed in Art. 6, Chap. X.) This circuit[1] is a particularly advantageous combination of two triodes that provides the gain and bandwidth of a single pentode, excellent stability, and a low

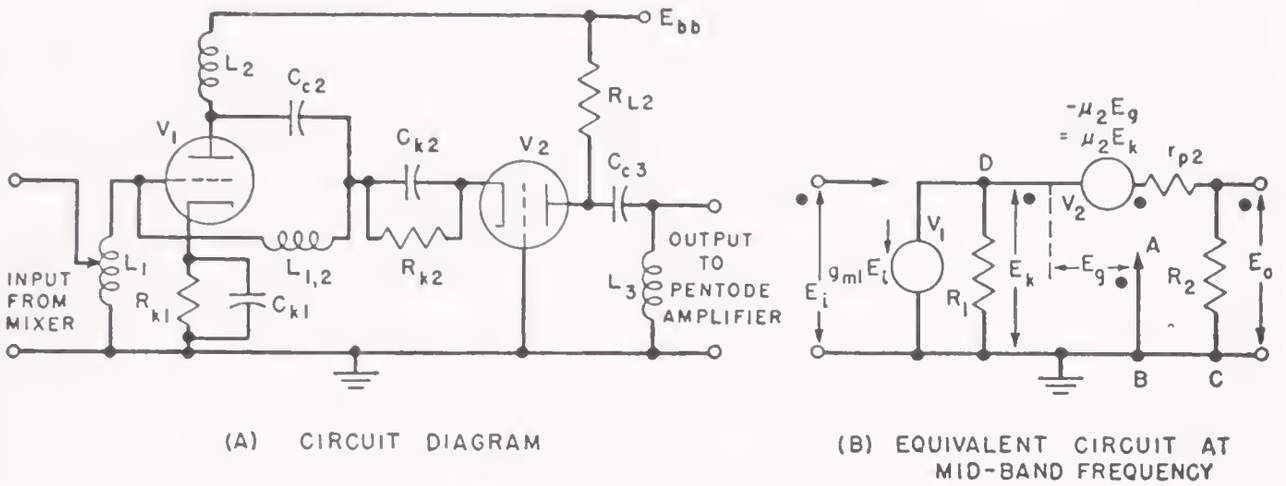


FIG. 43. Low-noise input stages for i-f amplifier.

noise figure essentially equal to that of the first triode. With a triode-connected 6AK5 as V_1 and a 6J4 as V_2 in a 30-Mcps i-f amplifier having a bandwidth of 6 Mcps, a noise figure of 1.35 may be attained. For a pentode 6AK5 employed as the first stage of a similar amplifier, the noise figure is about 2.1.

In Fig. 43*A* the input circuit and the coupling circuits are tuned to resonance at mid-band frequency by the inductors L_1, L_2, and L_3. Also inductor $L_{1,2}$ is adjusted to resonance with the grid-plate capacitance of V_1 at the same frequency. Thus, at the mid-band frequency, the equivalent circuit for Fig. 43*A* has the form indicated in Fig. 43*B*. Voltage E_i represents the voltage applied by the input circuit to the grid of V_1, E_k is the voltage applied by V_1 to the cathode of V_2, and E_o is the voltage delivered to the pentode stage that follows V_2. Resistor R_1 is the parallel combination of the plate resistance of V_1, the loss resistance of L_2, and the grid-cathode resistance of tube V_2 that represents grid-loading effects.

[1] H. WALLMAN, A. B. MACNEE, and C. P. GADSDEN, "A Low-noise Amplifier," *Proc. I.R.E.*, **6** (June, 1948), 700–708.

Similarly, R_2 represents the parallel combination of R_{L2}, the loss resistance of L_3, and the input resistance of the next stage. The plate load resistance for V_1 is the parallel combination of R_1 with the input resistance of the grounded-grid stage in Fig. 43*B*, and this input resistance is very low, of the order of a few hundred ohms. As a result, the voltage amplification of V_1 is approximately unity, and there is no tendency for V_1 to oscillate. Inductance $L_{1,2}$ is not needed to prevent oscillations, but it does provide some improvement of the noise figure.

Amplification and impedance values may be determined from the equivalent circuit of Fig. 43*B*. The voltage applied to r_{p2} and R_2 in series is $(\mu_2 + 1)E_k$. Therefore, the current drawn by these resistors is the same as would be drawn by a resistance $(r_{p2} + R_2)/(\mu_2 + 1)$ to which the voltage E_k is applied. Thus $(r_{p2} + R_2)/(\mu_2 + 1)$ is the low resistance faced by R_1. In wide-band amplifiers R_2 is much less than r_{p2}. The resistance faced by R_1 is therefore approximately $r_{p2}/\mu_2 = 1/g_{m2}$. Resistance R_1 is large and may be ignored in parallel with $1/g_{m2}$. Thus the voltage amplification of V_1 is very nearly g_{m1}/g_{m2}, a value that approximates unity if the two tubes are similar. Because R_1 is large compared with the resistance it faces, nearly the entire current $g_{m1}E_i$ is forced to flow through the R_2 branch of the circuit. That is, the grounded-grid stage acts to pass on to the load the current supplied to its cathode and supplies amplification in that the load voltage is greater than the cathode voltage. The over-all voltage amplification resulting from the current $g_{m1}E_i$ in resistor R_2 is $-g_{m1}R_2$. Thus the two triodes in Fig. 43 provide slightly more amplification in a given bandwidth than would a single pentode, the transconductance available in triodes being somewhat greater than that of pentodes.

The low noise figure of the grid-return amplifier in Fig. 43*A* may also be explained in terms of the equivalent circuit of Fig. 43*B*. The first step in this explanation is the proof of the following statement: If μ_2 in Fig. 43*B* is large relative to unity, the noise figure F_2 of the grid-return stage is negligibly different from that of a cathode-return amplifier employing the same circuit elements.

The voltage across the equivalent noise resistance of V_2 may be represented by a voltage source E_n in series with the connection AB in Fig. 43*B*. If the circuit is changed to cathode-return form (by connecting point C to point D instead of to point B), the output voltage produced by E_n is

$$E_o = \frac{-\mu_2 E_n R_2}{R_2 + r_{p2}} \tag{60}$$

If the grid-return connection is employed, the grid-cathode voltage is different from E_n because of the voltage drop caused by plate current in

R_1. Thus

$$E_g = E_n - \mu_2 E_g \frac{R_1}{R_1 + R_2 + r_{p2}}$$

or

$$E_g = \frac{E_n}{1 + \mu_2 R_1/(R_1 + R_2 + r_{p2})}$$

and the output voltage is

$$E_o' = \frac{-\mu_2 E_n R_2}{R_1(1 + \mu_2) + R_2 + r_{p2}} \tag{61}$$

Comparison of Eqs. 60 and 61 indicates that use of the grid-return connection multiplies the output voltage produced by E_n by the factor

$$\alpha_1 = \frac{R_2 + r_{p2}}{R_1(1 + \mu_2) + R_2 + r_{p2}}$$

A similar calculation made for a current source I_n connected between point D and ground shows that use of the grid-return circuit multiplies the output voltage produced by I_n by the factor

$$\alpha_2 = \frac{\mu_2 + 1}{\mu_2} \frac{R_2 + r_{p2}}{R_1(1 + \mu_2) + R_2 + r_{p2}}$$

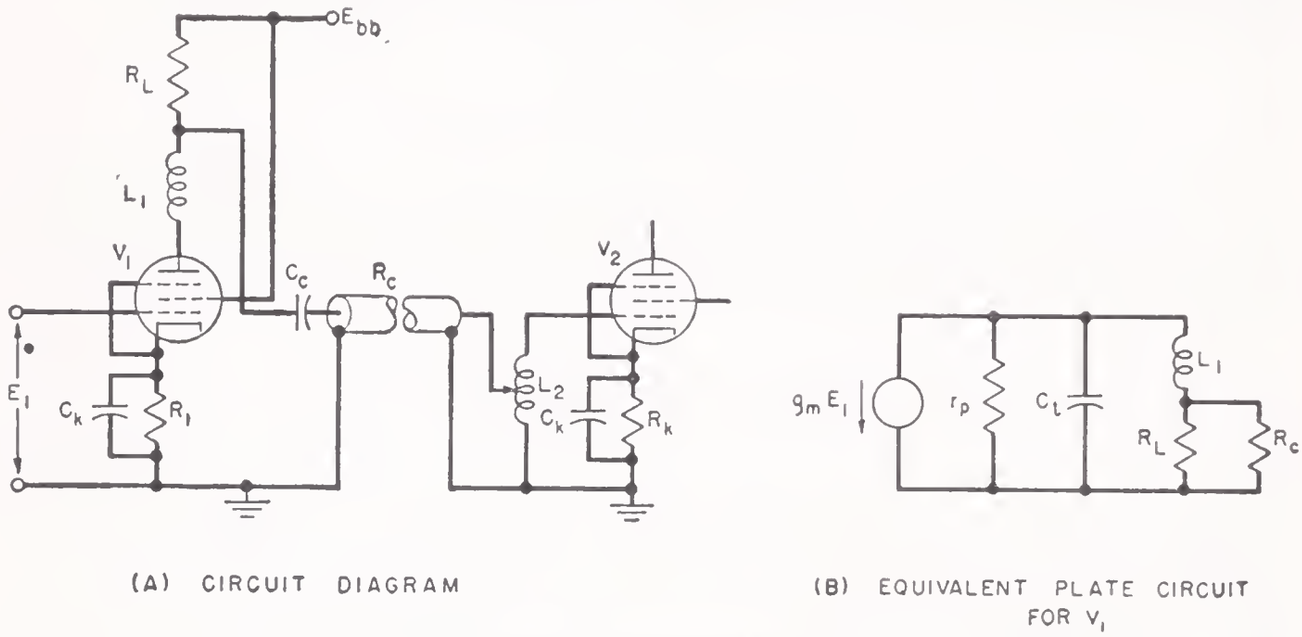
If $\mu_2 \gg 1$, the factors α_1 and α_2 are very nearly equal. Factor α_2 specifies the effect of the grid-return connection upon noise arising from the standard source in the noise-figure definition and upon noise arising in the grid circuit of tube V_1. Factor α_1 specifies the effect of the grid-return connection upon noise arising in the plate circuit of V_1 (represented by the equivalent noise resistance). To the extent that α_1 and α_2 are equal, the change to the grid-return connection modifies each component of noise to the same extent and does not change the noise figure.

Because the noise figure of V_2 is essentially the same as that of a cathode-return triode, it has its minimum value when the resistance of the source is somewhat less than the grid-loading resistance between grid and cathode of V_1 (see Art. 16). This resistance relation is provided by the circuit of Fig. 43*A*, the plate resistance of V_1 being somewhat less than the grid-loading resistance of V_2. Thus F_2 in the expression $F_1 + (F_2 - 1)/G_1$ for over-all noise figure is of the same order of magnitude as F_1, and the over-all noise figure is therefore very nearly equal to F_1.

The importance of noise considerations requires that, if a receiver contains no r-f amplifiers, the mixer and first i-f stages be located as close as possible to the antenna. Connecting lines and cables must not be allowed to attenuate the signal before it is amplified above the noise level. One procedure is to mount the entire i-f amplifier in a unit with the r-f

components. An alternate procedure is to mount the first two on three stages as an i-f preamplifier in the r-f unit and to connect these stages to the remainder of the i-f amplifier by means of an interunit cable.

Use of the preamplifier and connecting cable introduces the problem of matching the cable to the amplifier impedances. The cable would be equivalent to a large shunt capacitance and would decrease the bandwidth if it were terminated in the input impedance of a cathode-return amplifier stage, the input impedance being much greater than the characteristic resistance (about 50 ohms) of usual cables. Figure 44*A* shows the use of a tapped inductance at the output of a connecting cable. The inductance operates as a resonant transformer to change the high input



(A) CIRCUIT DIAGRAM (B) EQUIVALENT PLATE CIRCUIT FOR V_1

FIG. 44. Use of cable to connect stages of an i-f amplifier.

resistance of tube V_2 at mid-band frequency to the characteristic resistance of the cable. A similar tapped inductance could be employed at the input of the cable to provide V_1 with a sufficiently large load resistance.

An alternate impedance-adjusting scheme that is sometimes employed is indicated at the cable input in Fig. 44*A*. As shown by the equivalent circuit of Fig. 44*B*, the input resistance R_c of the matched cable appears in series with the tuning inductor L_1. (Resistor R_L is of the order of 1,000 ohms and may be ignored in parallel with R_c.) Capacitor C_t represents tube and wiring capacitance and resonates with L_1 in series with R_c. If X_L is the reactance of L_1 at resonance, the series combination of L_1 and R_c may be represented at the mid-band frequency as the parallel combination of an inductance with a resistance of value $R_c + X_L^2/R_c$. This resistance is the mid-band load resistance for V_1 and is much greater than R_c if X_L is large.

26. Diode Detectors. The detector of a radar receiver serves, as its position in the block diagrams of Fig. 1 (p. 339) indicates, to convert carrier-frequency pulses from the i-f amplifier into video pulses suitable,

after amplification, for application to the indicator tube. The simplest form of detector, and the form most commonly employed in radar receivers, uses a *diode* as a means of rectifying the intermediate-frequency pulses.

A typical circuit for a radar diode detector is indicated in Fig. 45*A*. Inductance L_i is the tuning inductance for the plate circuit of the final i-f-amplifier stage and is adjusted to resonate with tube and wiring capacitances. If an echo pulse of sufficient amplitude to make noise negligible is received by the radar system, the voltage e_i across L_i is an i-f pulse having an amplitude of the order of several volts and the form indicated by the dashed line in Fig. 45*B*. Positive excursions of e_i cause no current in the diode, but negative excursions result in a flow of current and a negative voltage e across R and C. Between the times of peak negative voltage of the e_i wave, capacitor C discharges through R. Thus

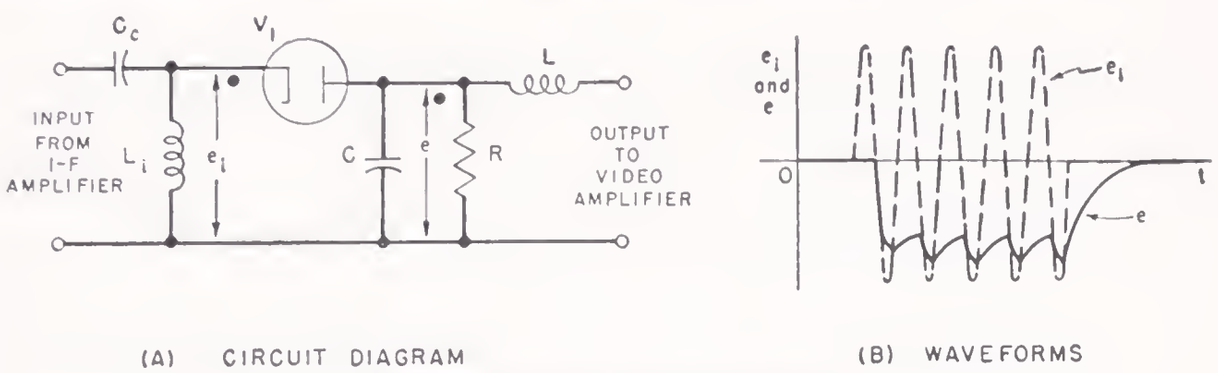


FIG. 45. Typical diode detector.

the waveform of e is a negative video pulse with sloping edges and superposed i-f ripple, as indicated in Fig. 45*B*. The negative polarity for the output pulse is ordinarily preferred, as explained in the discussion of video-amplifier blocking in Art. 19. If desired, however, positive pulses can be obtained by reversing the connections to the diode.

Inductance L in Fig. 45*A* in combination with wiring capacitance and the video-amplifier input capacitance (not shown in the figure) forms a low-pass filter designed to attenuate the i-f components in the waveform of e. The detector is ordinarily mounted on the i-f amplifier chassis, and the low-pass filter prevents appreciable i-f current from leaving this unit. In this way difficulties caused by feedback around the high-gain i-f amplifier are minimized. Attenuation of i-f components must be accomplished with the least possible loss of high-frequency components from the video signals at the detector output. Use of inductance L with small shunt capacitances provides much greater video bandwidth than would a single capacitor large enough to produce the same i-f attenuation. The filter employing L is comparable to a series-compensated video amplifier (Art. 21) in its effect upon the video-circuit frequency characteristic. The need for separating i-f and video-frequency components at

the detector output is one reason for employing relatively high intermediate frequencies in radar receivers (see Art. 18).

The input impedance of the diode detector has an important effect upon the output voltage developed by the final stage of the i-f amplifier. An equivalent circuit that indicates the general nature of this effect can be obtained if the following approximations are made: (1) The resistance (for positive plate voltage) and the interelectrode capacitance of the diode are assumed negligible. (2) The i-f voltage applied to the detector is assumed to be a perfect sine wave. (3) The ripple component in the detector output voltage is assumed negligible.

In accordance with the first assumption, the diode is represented in Fig. 46*A* as a switch that is operated so as to allow current to flow from the plate P to the cathode K, but not in the reverse direction. Current source $g_m E_g$ in Fig. 46*A* represents the current generator in the equivalent

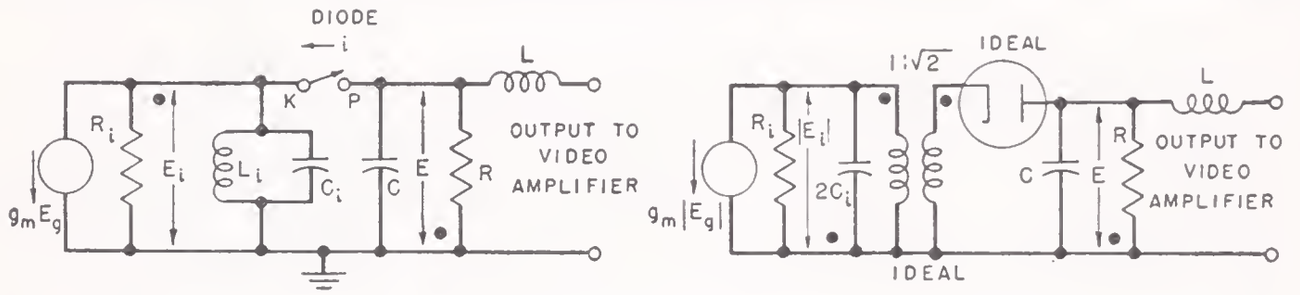


FIG. 46. Idealized equivalent circuits for diode detector.

circuit of the final i-f stage, and R_i and C_i represent the total shunt resistance and capacitance of the amplifier plate circuit and detector input circuit. Because the diode current i is a succession of very brief pulses, clipping of the peaks of the input-voltage wave E_i at the times of current flow occurs, and a ripple component appears in the output voltage E. Nevertheless, in accordance with the second and third assumptions, the input voltage is considered sinusoidal and the output voltage steady. The complex numbers E_g and E_i, representing sinusoidal voltages, and the real number E, representing a direct voltage, are considered to vary with time, but at rates so slow that the change during one i-f cycle is negligible.

In Fig. 46*A*, the magnitude of the steady voltage E must be equal to the peak value of the voltage E_i—closing of the switch prevents the peak value of E_i from exceeding E, and opening of the switch prevents the detector from maintaining a value of E greater than the peak input voltage. Thus

$$E = \sqrt{2}\,|E_i| \tag{62}$$

(The reference direction for E in Fig. 46*A* is chosen so that E is positive for the diode connections indicated.) If I and $|I_i|$ indicate, respectively,

the average value of the diode current i and the rms value of the fundamental component of i, the power output of the rectifier is EI and the power input is $|E_i|\,|I_i|$. Because the diode is assumed at all times to be either a short circuit or an open circuit, it is a lossless device, and $EI = |E_i|\,|I_i|$. Combining this relation with Eq. 62 yields

$$I = \frac{1}{\sqrt{2}}\,|I_i| \tag{63}$$

Equations 62 and 63 indicate that the detector behaves as an ideal transformer of step-up voltage ratio $1:\sqrt{2}$, except that it provides the additional feature of changing a-c rms values to d-c quantities. For example, if $|E_i|$ is the rms voltage applied to the detector, the output voltage is $\sqrt{2}\,|E_i|$ and the current delivered to a load resistance R_L in the d-c circuit is $\sqrt{2}\,|E_i|/R_L$. Under steady-state conditions the rms value of the fundamental component of the detector input current is then $\sqrt{2}\,(\sqrt{2}\,|E_i|/R_L)$, and the input resistance of the detector is

$$\frac{|E_i|}{|I_i|} = \frac{|E_i|}{\sqrt{2}\,(\sqrt{2}\,|E_i|/R_L)} = \frac{R_L}{2}$$

The ideal-transformer concept is made use of in the circuit of Fig. 46B, and this equivalent circuit is useful where an input wave of varying rms value is considered. The portion of the circuit to the left of the ideal transformer acts upon video signals equal at each instant to the rms value of the i-f waves in Fig. 46A. As explained in Art. 8, this circuit has the same effect upon the video signals as the actual circuit has upon the modulation envelope of the i-f signals, the capacitance $2C_i$ being substituted for the resonant circuit comprising L_i and C_i. The remainder of Fig. 46B represents the actual video circuit of the detector and its effect upon the video equivalent of the i-f circuit. The ideal diode is included to account for the effect observed at the trailing edge of the pulse in Fig. 45B. If the input voltage decreases more rapidly than video-circuit currents can discharge C, the diode becomes an open circuit, and i-f and video portions of the detector are disconnected.

The circuit of Fig. 46B shows that under conditions for which the diode does not become an open circuit the entire detector circuit may be considered a single interstage network. The rate of change of the rms value of the i-f voltage at the detector input is limited by the rate at which the video output voltage can be made to change. Calculations of response characteristics may be made if all elements of Fig. 46B are referred to the same winding of the transformer. For example, if L were open-circuited, the entire circuit would be analogous to a single-tuned i-f

stage or an uncompensated video amplifier, the bandwidth being determined by the time constant

$$\frac{2R_iR}{2R_i + R}(C_i + C)$$

Because of the idealizations upon which Fig. 46 is based, the equivalent circuit of Fig. 46*B* represents only approximately the performance of an actual detector. In particular, the amplitude of the output pulses is materially less than that predicted by the equivalent circuit. In typical radar circuits, the output amplitude may be of the order of half the value calculated from Fig. 46*B*. One reason for this loss is that to permit rapid response to pulse edges capacitor C must be made relatively small. Thus this capacitor discharges appreciably between successive peaks of the i-f wave, and the average value of the rectified voltage is reduced. A second reason is that the conducting resistance of the diode is not negligible but has a value of several hundred ohms. Because the diode current flows in brief pulses of large amplitude, diode resistance causes as large a voltage drop as a much larger series resistance in the output circuit. The d-c load resistance is ordinarily only a few thousand ohms, and therefore an important loss of voltage is caused by diode resistance. A third reason is that the interelectrode capacitance of the diode is not negligible relative to capacitor C. The diode capacitance and C form a voltage divider that applies to the diode only that fraction of the total i-f voltage that appears across the diode capacitance. A fourth effect appreciable in wide-band circuits is the loss of voltage occasioned by the peak-clipping effect of the diode current upon the resonant-circuit voltage.

In order to minimize the loss of voltage in a diode detector, the diode employed should be one having minimum resistance to current flow in the forward direction and minimum plate-to-cathode capacitance. Sometimes, in an effort to obtain desired resistance and capacitance values or as a means of reducing the number of tube types in a radar set, a triode or pentode tube is used as the detector, the grids being connected to the plate to form a diode. If a pentode has an internal connection between cathode and suppressor grid, the connections of Fig. 47 can be used to produce a low-capacitance diode. The control grid becomes the anode of the diode, and to remove the suppressor-grid-to-plate capacitance from the circuit, the plate is joined to the cathode. The screen grid shields the control

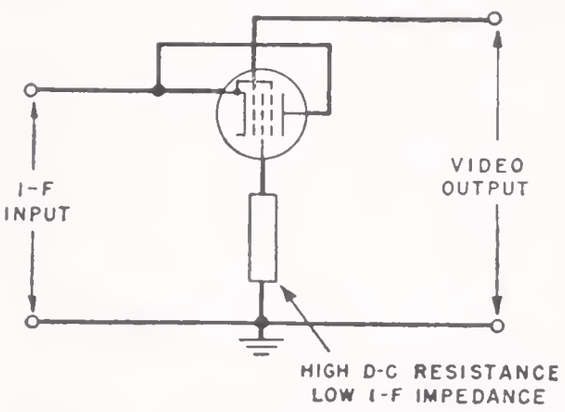


FIG. 47. Diode-detector connections for a pentode having internal connection between cathode and suppressor grid.

grid from the suppressor grid because it is at ground potential for i-f signals. The d-c resistance in the screen-grid circuit prevents the diode formed by cathode and screen grid from short-circuiting the detector input voltage.

The equivalent circuit of Fig. 46*B* emphasizes the respects in which the performance of a diode detector resembles that of a linear circuit. If the modulation envelope of the i-f wave is considered the input signal for the detector and the actual value of the output voltage is taken as the output signal, then under many conditions of operation the detector is representable by a linear circuit, and the principles of linear-circuit analysis may be applied to the entire i-f-amplifier, detector, and video-amplifier system. Because of the linear relation of output voltage to rms input voltage, the circuit is often called a *linear detector*.

A nonlinear element is required for rectification, nevertheless, and nonlinear effects appear in the performance of linear detectors under certain conditions. For example, the effect provided for by the diode in Fig. 46*B*, whereby the detector operates differently upon the leading and trailing edges of pulses, could not occur in a linear device. In addition, the relation of output voltage to rms input voltage is not linear in the region of very low-amplitude signals.

A nonlinear effect of great importance in radar receivers appears if two or more signals are applied to the detector simultaneously. Unless the carrier waves of the signals have exactly equal frequencies and phase differences of 0 or 180 deg, *the superposition principle cannot be applied to the detector*. The output obtained when two signals are applied simultaneously is *not* always the sum of the outputs obtained when the same signals are applied separately. If two sine waves of slightly different frequencies appear in combination at the detector input, the envelope of the total input wave is not the sum of the envelopes of the individual waves. Rather, the carrier waves swing into phase and out of phase because of the frequency difference, and a *heterodyning* or *beating* effect is produced. The total envelope is the sum of the individual envelopes when the waves are in phase and is the difference of the individual envelopes when a 180-deg phase displacement occurs. The total envelope varies between these limits at the *beat frequency*, which is the difference between the two carrier frequencies. The video output of the detector approximates the envelope of the total input wave and therefore contains the beat frequency.

Of particular interest in radar is the simultaneous application to the detector of weak echo-pulse signals and the noise that necessarily accompanies the signals. Because of the great amplification and limited bandwidth of the i-f stages, the noise voltages applied to the detector have tremendously greater amplitude than the noise voltages in the input cir-

cuits and are somewhat different in waveform. The detector-input noise has an amplitude of the order of several volts and may be thought of as an intermediate-frequency carrier wave that is subject to erratic variations of both amplitude and phase. This waveform, when applied to the detector in the absence of echo signals, produces a video waveform comprising erratic fluctuations in one direction from the zero axis. The d-c component of this waveform is, of course, blocked by the coupling capacitors in the video amplifier.

If an echo-pulse signal is added to the fluctuating noise voltage at the input terminals of the detector, the average output voltage of the detector is increased during the time of the pulse. The amount of the increase depends upon the amplitude of the signal in the manner indicated by the solid-line curve[1] of Fig. 48. Note that the output is less than that specified by the dot-dash line representing the superposition of signal and noise; the dot-dash line would apply if the detector were truly a linear device. The loss of output comes about because the phase angle between the noise and signal waves changes erratically through all values, and therefore the resultant input voltage fluctuates between the sum and the difference of the noise and signal voltages. A small signal subtracts nearly as much from the output caused by noise alone as it adds to it; therefore, the effect at the output caused by the signal is very small, and the signal is said to be *suppressed* by the noise. Increasing the signal amplitude at first causes the increment in the output voltage to grow in proportion to the square of the signal amplitude. For signal amplitudes greater than the rms noise level, the square law no longer applies, and for large signals, the noise has little effect upon the average value of the output voltage.

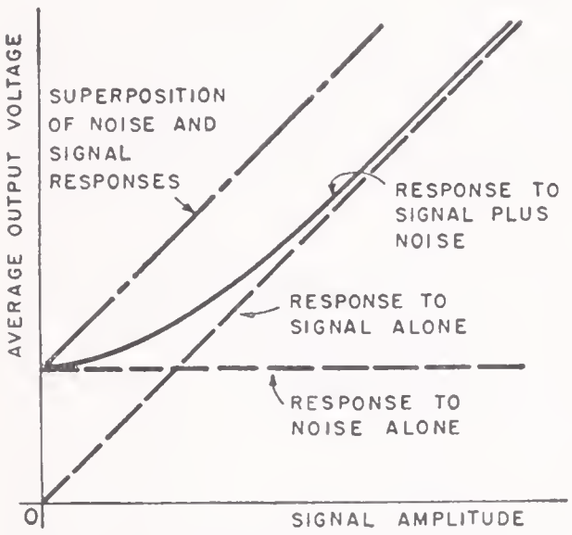


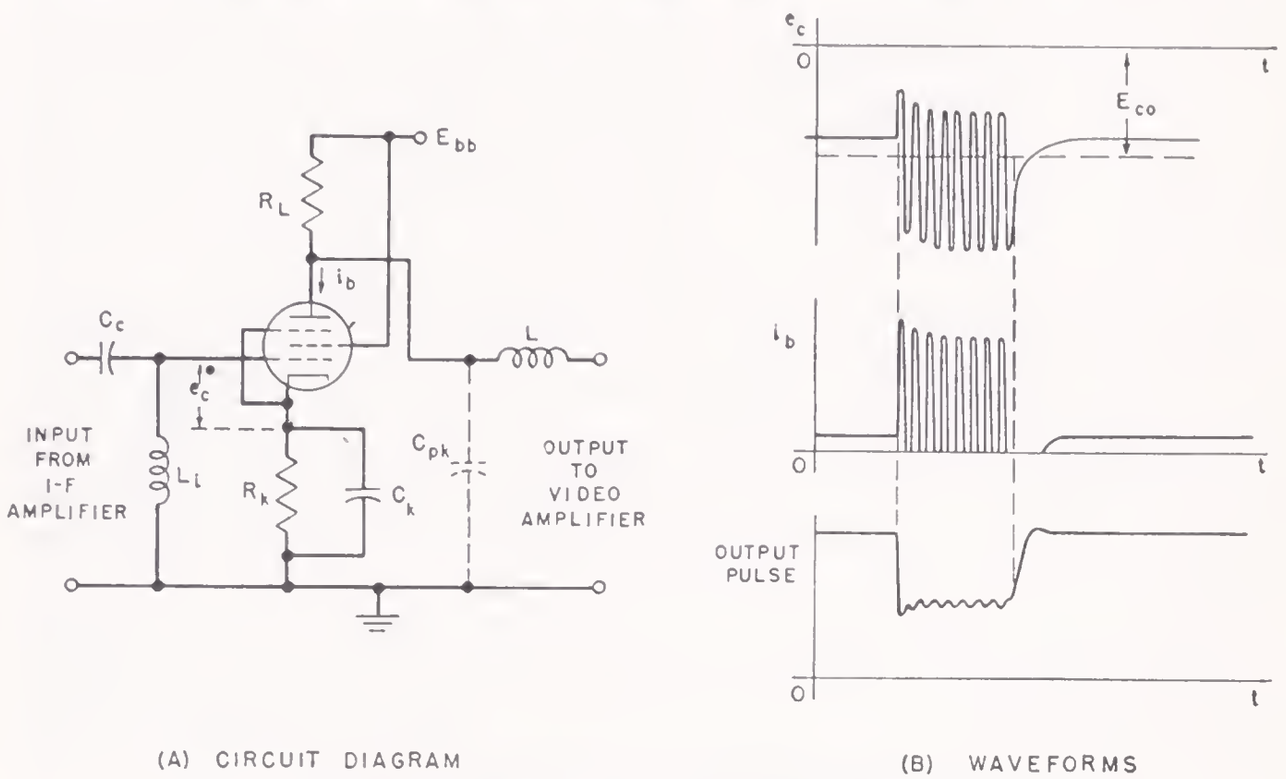
Fig. 48. Response of ideal linear detector to signal and noise.

The suppression of the signal by the noise at the detector is the reason that, as mentioned in Art. 18, the over-all receiver bandwidth should be controlled by the i-f amplifier rather than the video amplifier. A very wide video response band increases the noise amplitude somewhat but at the same time makes the signal pulses more noticeably different in shape from the noise peaks. Excess i-f bandwidth, on the other hand, increases the noise level at the detector and thus causes the suppression effect of the noise to extend to signals of larger amplitude.

[1] The mathematical form of this curve is derived in W. R. Bennett, "Response of Linear Rectifier to Signal and Noise," *B.S.T.J.*, **23** (January, 1944), 97–113.

The suppression effect also increases the difficulty of decreasing appreciably the minimum-detectable signal amplitude by means of special indicating devices. An improvement in the indication method that permits detecting video signals of one-fourth the original minimum amplitude merely halves the amplitude of detectable i-f signals because of the square-law effect of signal suppression by the noise.

27. Triode and Pentode Detectors. In some radar receivers, triodes or pentodes are employed as plate-circuit detectors, one circuit being of the form illustrated in Fig. 49*A*. The i-f signal is applied to the grid of the detector in series with a bias voltage of approximately the cutoff value.



(A) CIRCUIT DIAGRAM (B) WAVEFORMS

FIG. 49. Plate-loaded plate-circuit detector.

The plate current is therefore nearly zero between signal pulses and comprises, during the signal, a series of brief pulses of nearly half-sine shape (see Fig. 49*B*). The plate current causes a voltage drop across the load resistor R_L, and the video components of this voltage form a negative pulse that is the output voltage of the detector. The i-f ripple components are removed from the output voltage by a low-pass filter, as in the diode detector.

In Fig. 49*A*, a cathode self-bias circuit is indicated. An alternate possibility, that of returning L_i to a negative voltage source, has the disadvantage that a low-resistance source of the bias voltage would be required. Otherwise the discharge time constant of C_c would be long, and the detector would be blocked for an appreciable period after the application of a signal pulse of sufficient amplitude to cause grid-current flow. If the capacitance C_k in the self-bias circuit is made large enough to maintain constant voltage during the radar-set repetition period, the

bias voltage provided will depend upon the number and amplitude of the signal pulses present, and no adjustment of R_k will yield the correct bias voltage under all signal conditions. A better arrangement is to make C_k large enough to serve as an i-f by-pass, yet small enough to permit rapid change of the bias voltage at the pulse edges. Resistor R_k is then adjusted to provide a very small value of no-signal plate current—complete cutoff of plate current is impossible because current in R_k is needed to produce the bias voltage. The disadvantage of the small by-pass capacitor is that the detector gain is reduced because (as indicated in Fig. 49*B*) the changes in bias voltage partially offset the effect of the signal in changing the plate current.

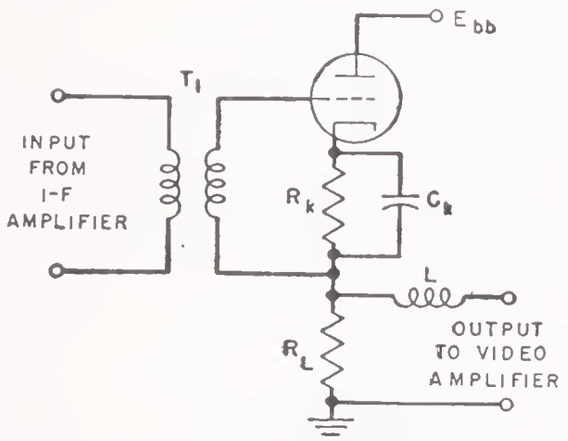


FIG. 50. Cathode-loaded plate-circuit detector.

If positive output pulses are desired, rather than the negative pulses obtained from the circuit of Fig. 49, the load resistance may be moved to the cathode circuit, as in Fig. 50. If a double-tuned coupling circuit is employed at the detector input in the manner illustrated, the grid-cathode voltage is not influenced by the output voltage, and the cathode-loaded and plate-loaded detectors operate in essentially the same manner.

An alternate connection of the cathode-loaded circuit, which resembles a cathode follower and is sometimes called an *infinite-impedance detector*, is illustrated in Fig. 51. Operation of this circuit differs from that of Fig. 50 because the cathode-to-ground voltage E_k appears in the grid circuit as a negative bias voltage that very nearly cancels the effect of the input signal in increasing the plate current. This feedback effect is somewhat similar to that in a cathode follower. The circuit is not a cathode follower, however, because C is an i-f by-pass capacitor, and the output voltage returned to the grid is a video pulse, whereas the input is an i-f signal.

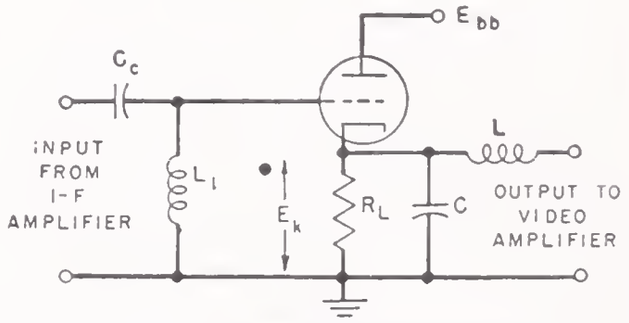


FIG. 51. Cathode-follower connection of plate-circuit detector.

The detector of Fig. 51 is perhaps best thought of as equivalent to a diode detector, except that the output current is obtained from the plate power supply rather than from the source of the i-f signals. The plate-cathode path in the tube may be thought of as a switch that is closed

whenever the peak value of the input voltage exceeds $E_k - E_{co}$, E_{co} being the cutoff bias voltage.

Plate-circuit detectors have the obvious advantage relative to diode detectors of providing greater gain; that is, for signal amplitudes within the range of linear-detector operation, the video output obtained with a specified i-f input is greater. In addition, greater gain can be had from the final stage of the i-f amplifier when plate-circuit detectors are employed, because the detector output power is obtained from the plate-voltage supply rather than from the i-f amplifier.

The major disadvantage of plate-circuit detectors is the small range of linear operation. The video output of diode detectors is very closely proportional to the peak value of the input wave for all signal amplitudes greater than about 1 volt. The linear range of plate-circuit detectors, on the other hand, has (in addition to the small-signal limit) a large-signal limit caused by the flow of grid current. The linear range is therefore undesirably small, particularly for the circuits of Figs. 49 and 50. Another disadvantage of triode and pentode detectors is the increased probability that incidental effects—such as power-supply hum and feedback—will interfere with the desired operation of the circuit.

28. Crystal Mixers. The name *mixer* refers to a nonlinear device in a superheterodyne receiver in which a continuous local-oscillator wave and a modulated r-f signal wave are combined to produce an intermediate-frequency wave carrying the same modulation as the r-f wave. Low-frequency radio receivers commonly employ multigrid tubes that function as both mixers and local oscillators and are called *frequency converters*. These tubes do not operate satisfactorily at radar frequencies; therefore, radar receivers use separate mixers and local oscillators, crystal mixers being employed for microwave reception and vacuum tubes (usually triodes) being employed most frequently in uhf systems (see Fig. 1, p. 339).

The r-f circuits of microwave crystal mixers make use of transmission lines, waveguides, and cavity resonators. These high-frequency circuit elements are treated in Chaps. VII to IX, and discussion of the actual forms of microwave mixers is therefore reserved for Chap. XII (see Arts. 16 and 17). The present article is concerned with the principles applicable to lumped-parameter equivalent circuits of mixers and is a necessary preliminary to the discussions in Chap. XII.

A basic equivalent circuit in terms of which the major features of mixer operation can be explained appears in Fig. 52*A*. The signal source (I_s in parallel with resistor R_s) and the local oscillator send current through the tuned circuit comprising L_s and C_s. Thus the mixer input voltage e is, as shown in Fig. 52*B*, the sum of a continuous sine wave e_{osc} of local-oscillator frequency and a pulse-modulated wave e_s of signal frequency. The local

oscillator is represented in Fig. 52*A* as a current source because, to prevent loss of signal power to the oscillator, it is coupled to the mixer through a high impedance. The L_s, C_s circuit is tuned to the frequency of I_s in order that the largest possible signal voltage may be applied to the mixer. Nevertheless the local-oscillator component of e is ordinarily enormous in comparison with the signal voltage; the oscillator voltage is of the order of

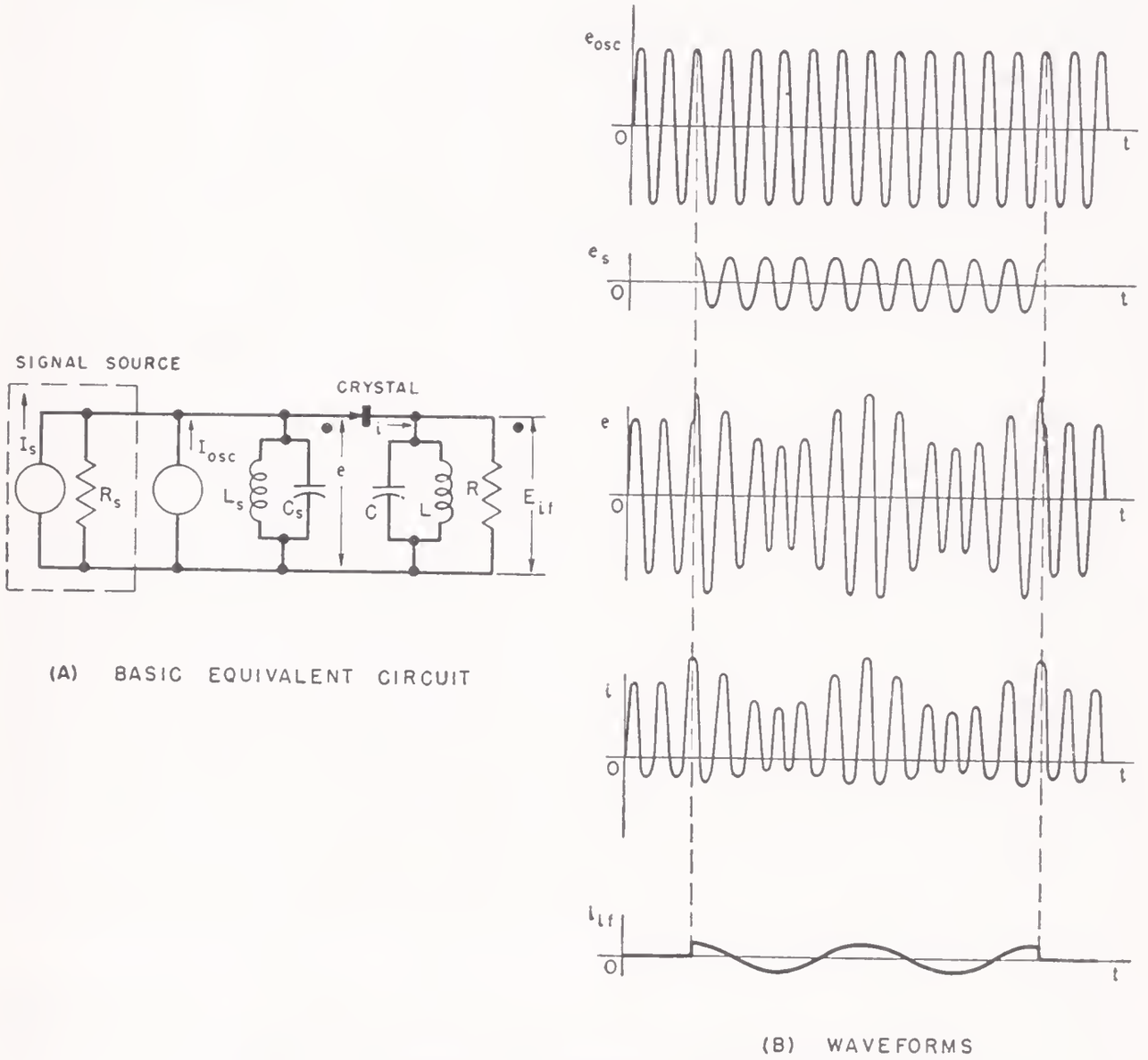


FIG. 52. Diagrams for explaining mixer principles.

a few tenths of a volt, whereas the signal voltage may be less than a microvolt.

Because the crystal conducts current more readily in one direction than the other, the crystal current i is an unsymmetrical wave, as indicated in Fig. 52*B*. During periods between the application of signal pulses, i comprises a d-c component, a fundamental component at the oscillator frequency, and higher harmonic components at multiples of the oscillator frequency. When a signal pulse is applied, additional frequency components appear. In particular, because the average value (over a few r-f cycles) of i increases and decreases in correspondence with the beat-

frequency envelope of e, a current component at the difference frequency of the oscillator and signal waves appears. By proper adjustment of the oscillator frequency, the difference frequency can be made to equal the desired intermediate frequency of the receiver. Thus the intermediate-frequency component i_{if} (see Fig. 52*B*) of the crystal current i is produced by the mixer.

The output voltage of the mixer results from the passage of the current i through the parallel combination of L, C, and R (see Fig. 52*A*). Because the output circuit is tuned to the intermediate frequency, only the component i_{if} produces an appreciable voltage drop, and the output voltage is an i-f pulse having the same modulation envelope as the r-f input pulse.

The foregoing explanation suggests that a crystal mixer is in principle a diode detector, the intermediate frequency being produced by the beat effect of two simultaneously applied signals. Note, however, that in order to emphasize the desired output-voltage component the load impedance for the mixer is made negligible except at the intermediate frequency. For this reason essentially no output voltage is obtained when local-oscillator voltage alone is applied, and the conduction period of the nonlinear mixer element endures for roughly half of each local-oscillator cycle.

Of particular significance is the following contrast between detectors and mixers: As explained in Art. 26, the detector is fundamentally a *nonlinear* device. The mixer, on the other hand, when operated with small input signals, may be thought of as a *variable-parameter linear* device. To be sure, the crystal in the mixer is actually nonlinear. Nevertheless, because of the small amplitude of the signal voltage relative to the local-oscillator voltage, it may be assumed that the incremental conductance or slope of the i, e characteristic of the nonlinear element is controlled as a definite function of time by the oscillator voltage and is negligibly altered by the small signal voltage. With respect to signal quantities, therefore, the mixer is characterized by a parameter that varies as a fixed function of time. Nonlinear effects can occur only if the applied signals are so large that the mixer parameters become functions of the signal voltage as well as the oscillator voltage. As long as only small signals are applied, the principle of superposition may be applied to the mixer, special effects such as the suppression of signals by noise in detectors cannot occur, and a mixer behaves in every respect as an ordinary linear network—except, of course, a change of frequency is produced because of the variable element.

The linear-network characteristics of a mixer are emphasized in the equivalent circuit drawn in Fig. 53. The crystal of Fig. 52*A* is replaced by a linear network, and the local-oscillator current I_{osc} is omitted because the linear equivalent applies only to signal quantities. The

parameters shown in the network representing the crystal are the short-circuit parameters. Suppose, for example, that a short circuit is placed at the output terminals and an r-f signal voltage is applied at the input terminals. The resulting r-f component of crystal current gives rise to

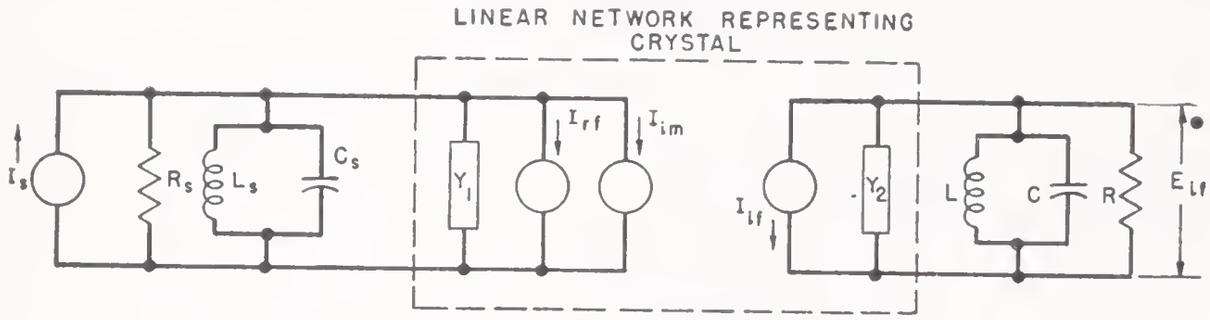


FIG. 53. Equivalent signal circuit for crystal mixer.

the short-circuit input admittance represented by Y_1, and the corresponding i-f component is the short-circuit output current of the mixer represented by I_{if}. The output voltage and current for any value of load admittance may be determined if the mixer output circuit is considered to be a current source I_{if} shunted by the output admittance Y_2. Admittance Y_2 is the admittance presented to an intermediate-frequency voltage source at the output terminals when the input terminals are short-circuited. The i-f source at the output causes high-frequency currents in the input short circuit. These currents, represented by I_{rf} and I_{im}, may be explained as follows:

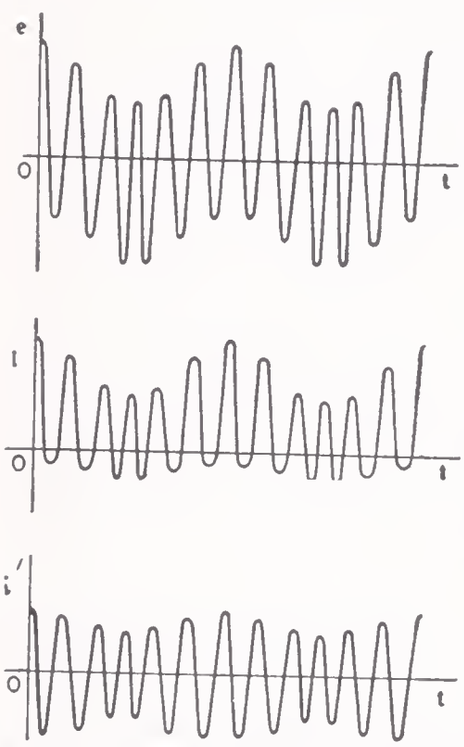


FIG. 54. Mixer waveforms for i-f signal applied at output.

For the conditions considered, the voltage at the nonlinear element, which is the sum of the high-frequency local-oscillator voltage and an intermediate-frequency voltage of small amplitude, may be represented by the wave marked e in Fig. 54. The corresponding crystal-current wave is represented by i, and this wave with d-c and i-f components removed is drawn as i'. Observe that i' has the form of a sinusoidally modulated carrier wave and may therefore be considered the sum of a carrier wave of local-oscillator frequency and two side-band waves at the oscillator frequency plus and minus the intermediate frequency. One of these side-band current waves is at the frequency of I_s in Fig. 53 and is represented by I_{rf} in Fig. 53. The other is at the image frequency (see Art. 18) and is represented by I_{im}.

In a complete analysis of a mixer, it is necessary to consider not only the currents of signal frequency, image frequency, and intermediate fre-

quency indicated in Fig. 53 but also current components of many other high frequencies. All components except the three indicated in Fig. 53 have only small effects upon the operation of mixers of the usual design and are omitted from consideration here. Note, however, that the components produced by simultaneous application of oscillator and r-f-signal voltages have frequencies equal to the differences between the signal frequency and harmonics of the oscillator frequency. It is thus possible to use a local oscillator that operates at approximately a submultiple (usually half) of the signal frequency. Such systems have been used in radar receivers but are avoided whenever an oscillator of the desired frequency range is available, because harmonic operation increases the loss of signal in the mixer.

Admittances Y_1 and Y_2 represent the r-f and i-f admittances of the mixer only under the short-circuit condition. The r-f input admittance is in general (see Fig. 53) the ratio of the sum of I_{rf} and the current in Y_1 to the applied r-f voltage. Current I_{rf} depends upon the i-f voltage across the load impedance, and the r-f input admittance is therefore a function of the i-f load impedance. In a similar fashion, source I_{if} depends upon r-f signal voltages, and the i-f impedance of the mixer is a function of the input-circuit impedance at the signal frequency.

Current source I_{if} depends, in fact, not only upon voltages at the r-f signal frequency arising from I_{rf} but also upon voltages at the image frequency arising from I_{im}. Input circuit impedances at both the signal and image frequencies therefore influence the i-f output admittance of the mixer. Image-frequency impedances also influence slightly the r-f input impedance, because changes produced in the i-f circuit in turn influence the r-f circuit. The r-f circuit does not, in general, present the same impedance to the mixer input terminals at both signal and image frequencies. The input circuit in Fig. 53 may be tuned and matched at the signal frequency, so that maximum power is transferred from the signal source to the mixer. At the image frequency, however, the input circuit is not resonant, and the input-circuit impedance may be nearly a short circuit. Independent control of impedance relations at the two frequencies is required if adjustment for the optimum conditions of operation is to be made.

The effectiveness of a mixer in converting r-f power to i-f power is usually specified in terms of the *conversion loss* of the mixer. By conversion loss is meant the ratio of the *available* power (see Art. 14) of the r-f source to the *available* i-f output power of the mixer; thus conversion loss is the reciprocal of the available-power gain (see Art. 17) of the mixer. For an ideal mixer element consisting of a lossless switch, the output i-f power would equal the r-f input power, provided no power were lost from the mixer at the image frequency or at harmonic frequencies, and for a tuned

and matched input circuit, the conversion loss would be unity (or zero decibels). In an actual mixer, a conversion loss of 6 db, corresponding to an output power equal to one-fourth the input power, is typical. This loss occurs even though the input circuit is tuned and matched, and power loss at image and harmonic frequencies is negligible. The loss may be ascribed to the fact that the crystal is not an ideal switch but has the equivalent circuit of Fig. 19, p. 373. In particular, large r-f currents in capacitor C of Fig. 19 must flow through the spreading resistor R. Values of conversion loss obtained in practice depend, of course, upon the crystal employed, upon the input-circuit impedance relations at the signal and image frequencies, and upon the local-oscillator power. Values for standard crystals under typical operating conditions are listed in Table 1, p. 376.

Because of the very small amplitude of the signal at the mixer, not only the loss of signal power but also the introduction of noise in the mixer are important to over-all receiver performance. As explained in Art. 16, the noise generated in a crystal may be specified by a noise-temperature ratio t. Noise may be accounted for in Fig. 53 by the addition of a noise-current source in parallel with I_{if}, the i-f noise power available from this current source in parallel with the output admittance of the mixer being t times the available noise power ($kT\Delta f$) of a linear, passive network. The value of t depends upon the local-oscillator power and to a very small extent upon impedance adjustments in the r-f circuit. For normal operating conditions, values ranging from 2 to 3 are typical; specific values are listed in Table 1.

The definitions of t, the noise-temperature ratio, and of F_{mix}, the noise figure (see Art. 17) of the mixer, may be written in terms of ΔN, the available i-f noise power in the frequency band Δf, and G_{mix}, the available-power gain of the mixer (the reciprocal of the conversion loss), as

$$t = \frac{\Delta N}{kT\Delta f}, \qquad F_{mix} = \frac{\Delta N}{G_{mix}kT\Delta f}$$

Therefore,

$$F_{mix} = \frac{t}{G_{mix}} \tag{64}$$

The over-all noise figure of the mixer and i-f amplifier is, by Eq. 35, p. 394,

$$F = F_{mix} + \frac{F_{if} - 1}{G_{mix}} \tag{65}$$

where F_{if} is the noise figure of the i-f amplifier. Combining Eqs. 64 and 65 yields

$$F = \frac{t + F_{if} - 1}{G_{mix}} \tag{66}$$

The noise figure F calculated from Eq. 66 is somewhat less than the over-all noise figure actually obtainable in a microwave receiver for two reasons: (1) The r-f signal power available at the mixer input is less than the power available from the antenna because of losses in the input system. (2) The noise at the mixer output is increased by the presence of noise originating in the local oscillator. The oscillator supplies to the mixer not only a desired sine wave of large amplitude but also a noise wave of small amplitude having frequency components in a fairly broad band surrounding the oscillator frequency. Noise components at the signal and image frequencies are converted in the mixer to intermediate-frequency noise. The effects of input-circuit loss and local-oscillator noise can be taken into account by rewriting Eq. 66 as

$$F_{rec} = \frac{t + \Delta t + F_{if} - 1}{G_{mix} G_{rf}} \tag{67}$$

where F_{rec} is the over-all noise figure of the receiver, G_{rf} is the available-power gain (number less than 1) of the r-f input circuits, and Δt is an apparent increase in the mixer-noise-temperature ratio caused by local-oscillator noise.

Equation 67 can be used not only for the calculation of over-all noise figures but also for determining the most desirable operating conditions of the mixer. Consider, for example, the effect of impedance adjustments in the mixer circuit upon the quantities in the right-hand member of Eq. 67. The input-circuit gain G_{rf} is, of course, independent of mixer adjustments, and $t + \Delta t$ is very little changed when mixer impedances are altered. The conversion loss $1/G_{mix}$, however, depends upon input-circuit impedance relations and is least if the input circuit is adjusted so that a tuned and matched i-f load produces the tuned and matched condition in the input circuit. The tuned and matched i-f load is used only to permit correct adjustment of the input circuit. Under operating conditions, a higher value of i-f load resistance is employed in order to minimize F_{if} (see Arts. 17 and 25). Thus a mismatch occurs also in the r-f circuit under operating conditions, but the r-f mismatch is considerably less than the i-f mismatch because of the conversion loss of the mixer.

The image-frequency impedance of the input circuit must be considered in making mixer impedance adjustments. Image-frequency adjustments have some effect upon G_{mix}; in particular G_{mix} is reduced if the mixer faces a resistive impedance at the image frequency and therefore delivers power at this frequency. Most important, however, is the effect of image-frequency impedance upon the i-f output impedance. Consistent results of i-f adjustments are possible only if the image-frequency impedance is carefully controlled.

Because of the difficulty of making impedance adjustments in mixers,

standard-type crystals are required to be reasonably uniform in impedance characteristics. Thus correct r-f and i-f impedance relations can be established in the design of a mixer, and no adjustments during use are required. For the crystals listed in Table 1, tuned-and-matched conditions are obtained if the r-f source resistance is about 50 ohms and the i-f load resistance is of the order of 500 ohms. (Type 1N28 crystals require a considerably higher load resistance.)

Another mixer adjustment that affects the quantities in Eq. 67 is the amplitude of the local-oscillator voltage applied to the crystal. A convenient measure of the oscillator voltage is the direct current produced in the mixer by rectification of the oscillator voltage. Provision is made in radar receivers for measuring this current, and in Fig. 55 typical curves of conversion loss and noise-temperature ratio are plotted as functions of the crystal current. From these curves and from Eq. 67, it may be seen that high noise figures result at low crystal currents because of the high conversion loss (low G_{mix}) and again at high crystal currents because of the high noise-temperature ratio $t + \Delta t$. The crystal current for minimum F_{rec} depends somewhat upon the crystal and the noise figure of the i-f amplifier but is usually approximately 0.5 ma. Appreciable change of crystal current from the optimum value can occur with only a small increase of the noise figure.

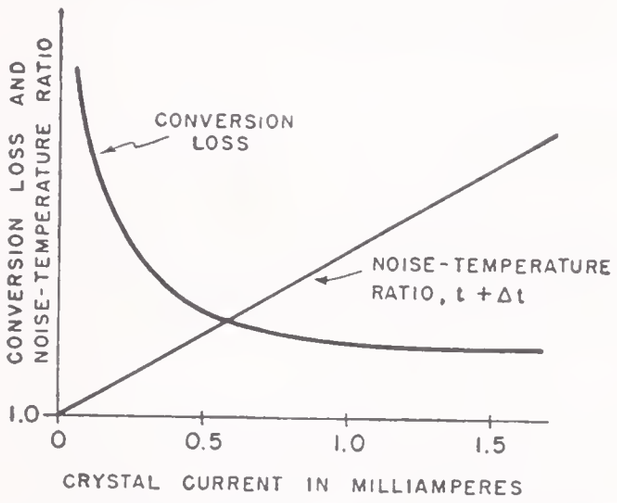


FIG. 55. Dependence of conversion loss and noise-temperature ratio upon crystal current.

The possibility exists of applying to the crystal a direct voltage in series with the local-oscillator voltage. A voltage in the direction to decrease the conduction period and the rectified current increases both the noise-temperature ratio and the conversion loss and should be avoided. For this reason, the d-c resistance of the crystal circuit must be kept less than about 100 ohms. A bias of a few tenths of a volt in the direction to increase the conduction period produces a slight improvement in noise figure and provides minimum noise figure over a somewhat wider range of values of crystal current. The improvement thus obtained, however, is seldom sufficient to justify the additional circuit complexity.

The effect of local-oscillator noise, represented by Δt in Eq. 67, upon the over-all noise figure is unimportant in 10-cm receivers but results in noise figure increases of about 2 and 4 db, respectively, in 3- and 1-cm receivers. The increased importance of oscillator noise in high-frequency receivers occurs because the bandwidth of the oscillator noise spectrum is

roughly proportional to the oscillator frequency. If higher intermediate frequencies are employed in the higher frequency receivers to avoid the increase of oscillator noise, a greater i-f amplifier noise figure results. Many 3- and 1-cm receivers employ *balanced mixers* to avoid the effect of local-oscillator noise. These mixers, described in Art. 17, Chap. XII, employ two crystals connected in such a way that the oscillator-noise outputs of the crystals cancel.

29. Vacuum-tube Mixers. As indicated in Fig. 1*A*, p. 339, uhf radar receivers ordinarily employ r-f amplifiers and vacuum-tube mixers. At frequencies up to about 500 Mcps, r-f amplifiers can be made with noise figures equal to or lower than those of the crystal-mixer input circuit.

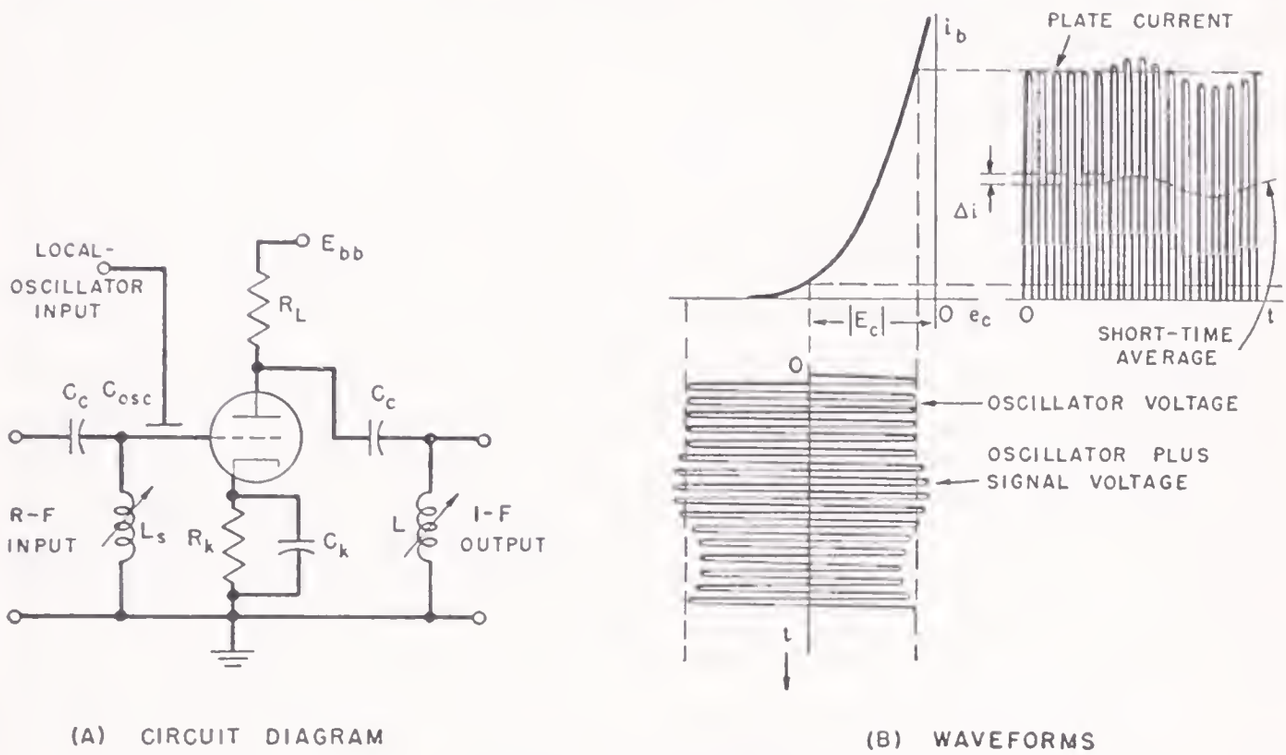


FIG. 56. Triode mixer.

For use with an r-f amplifier, a vacuum-tube mixer is preferred to a crystal mixer because of its greater reliability and conversion gain—the poorer noise figure of the vacuum-tube mixer is of no consequence if the signal is sufficiently amplified before application to the mixer.

Tubes used as radar mixers are special uhf types designed for operation at frequencies of 100 to 1,000 Mcps. Triodes similar to those employed in uhf oscillators (see Art. 1, Chap. X) and r-f amplifiers (see Art. 15, Chap. XII) are ordinarily used. Diodes can operate as mixers in essentially the same manner as crystals, but they do not provide the amplification available from triodes. Tetrodes and pentodes can be employed as mixers, but they are not so readily available in uhf designs as are triodes.

A typical triode mixer circuit for use at frequencies of 200 Mcps or less is indicated in Fig. 56*A*. Mixer circuits for the higher uhf range are similar in principle but employ transmission-line r-f tuning elements (see

Art. 15, Chap. XII). The tuning inductance L_s in Fig. 56*A* comprises three or four turns of heavy wire and is adjusted to resonate with tube and wiring capacitances at the frequency of the signals obtained from the r-f amplifier. The local-oscillator voltage enters the mixer circuit through C_{osc}, a very small coupling capacitance provided by a probe or disk connected to the oscillator output terminal and located near the mixer grid lead. Because of nonlinearities in the tube characteristics, a difference-frequency component is produced in the plate current when signal and oscillator voltages are applied to the grid simultaneously. For the correct local-oscillator adjustment, the difference frequency is equal to the intermediate frequency of the receiver, and an intermediate-frequency voltage appears at the output terminals—inductance L resonates with the tube and wiring capacitances at the intermediate frequency.

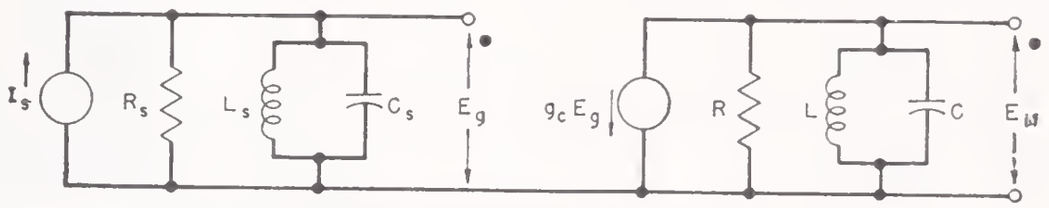


FIG. 57. Equivalent signal circuit for triode mixer.

Production of the i-f current component occurs in very nearly the same manner as in the crystal mixer and may be explained by reference to Fig. 56*B*. The grid voltage is the sum of a steady bias voltage E_c and sinusoidal oscillator and signal voltages. As indicated in the figure, beat-frequency variations of the envelope of the grid-voltage wave occur when oscillator and signal voltages are applied simultaneously. Corresponding plate-current variations may be determined by reference to the i_b, e_c curve drawn. Because the bias voltage is approximately the value for cutoff, plate current flows in essentially half-cycle pulses, and the average value (over a few r-f cycles) varies in accordance with the beat-frequency envelope when a signal is applied. Observe that, because of the presence of the local-oscillator voltage, there is no difficulty in obtaining essentially cutoff bias by means of a cathode resistor.

The triode mixer, like the crystal mixer, can be looked upon as a variable-parameter linear device rather than as a nonlinear device. In normal, small-signal operation, g_m, the transconductance of the tube, varies as a definite function of time in accord with the oscillator-voltage variations and is unaffected by the small signal voltage. Thus the triode mixer can be represented by the equivalent signal circuit of Fig. 57. The r-f amplifier tube from which input signals are obtained may be considered a current source in shunt with a resistor and a capacitor. The current source is represented by I_s in Fig. 57, and the resistor, in parallel combination with the loss resistance of L_s and the input resistance of the mixer tube, is represented by R_s. The amplifier capacitance, the mixer

input capacitance, the capacitance of the local-oscillator circuit, and the shunt capacitance of the input coupling circuit are represented by C_s. Parameters of the mixer output and i-f amplifier input circuits are represented similarly by R, L, and C, and current source $g_c E_g$ represents the i-f component of the current delivered by the mixer to a short-circuit load. Because of the linear properties of the mixer, $g_c E_g$ is proportional to E_g. Thus g_c, the *conversion transconductance* of the mixer, is a constant analogous to g_m but applicable to the frequency-conversion process.

The value of g_c depends upon the shape of the i_b, e_c curve in Fig. 56*B* and upon the bias and oscillator voltages applied to the tube. Maximum g_c is obtained if g_m is zero during one half of the oscillator cycle and as large as possible during the other half cycle—in this way the maximum change in the average plate current is obtained when the signal voltage is added to the oscillator voltage. For maximum g_c, therefore, the bias voltage is adjusted to the cutoff value, and the oscillator voltage is made as large as possible without the flow of grid current. Grid current is undesirable because of the loading of the signal source that it would produce. If the i_b, e_c curve were a straight line of slope g_m, and if cutoff bias were used, the increment of plate current caused by the application of a signal voltage in phase with the oscillator voltage and of peak value E_{sm} would be a half cycle of a sine wave of peak value $g_m E_{sm}$ and average value $g_m E_{sm}/\pi$ (see Eq. 2, p. 342). This average-value increment corresponds to the current change marked Δi in Fig. 56*B* and therefore to a value of g_c equal to g_m/π, or $0.318 g_m$. Actual tubes do not, of course, have straight i_b, e_c curves; under optimum operating conditions a value of g_c about 0.28 times the value of g_m at the peak of the oscillator-voltage cycle may be expected.

The equivalent circuit of Fig. 57 suggests that for the triode mixer, in contrast to a diode or crystal mixer, i-f impedance adjustments in the output circuit do not affect the r-f impedance of the input circuit, and vice versa. Complete independence of r-f and i-f impedances is obtained only if the output tuned circuit is a short circuit at the input frequency, and the input tuned circuit a short circuit at the intermediate frequency, so that no r-f voltage appears in the output circuit or i-f voltage in the input circuit. The short-circuit conditions are not completely realized in practice, and therefore some interaction between r-f and i-f circuits does exist. The interaction is small, however, so that a triode mixer, unlike a triode amplifier, can supply appreciable voltage amplification at high frequency without danger of oscillation resulting from feedback through the grid-plate capacitance.

Because the triode mixer of Fig. 57 is ordinarily preceded by r-f amplifier stages, noise generated in the mixer is relatively unimportant, and gain and bandwidth considerations are of major interest in the design of

the mixer. Nevertheless, an estimate of the noise generation of the mixer is needed to determine the amount of r-f amplification required to make the mixer noise negligible. As explained in Art. 16, noise generated in a triode comprises induced-grid noise associated with the transit-time grid-loading resistance and reduced-shot-effect noise representable in terms of an equivalent noise resistance R_{eq} in the grid circuit. For a tube used as a mixer, the grid-loading resistance is somewhat increased because plate-current flows for only half the oscillator cycle, and the importance of induced-grid noise is correspondingly decreased. Nevertheless, the noise figure is materially higher than for the same tube used as an amplifier because of the considerably higher value of R_{eq} in the mixer application. Where noise performance is of importance, therefore, a vacuum-tube mixer should not be used as a receiver input stage.

The higher value of R_{eq} that applies when a tube is employed as a mixer rather than as an amplifier may be explained as follows: By the definition of R_{eq} made in connection with Eq. 29, p. 389, the value of R_{eq} is proportional to $\Delta|I|^2/g_m{}^2$ for an amplifier tube and to $\Delta|I|^2_{mix}/g_c{}^2$ for a mixer tube, $\Delta|I|^2$ being the mean-square value of the plate-circuit noise current for the amplifier tube given by Eq. 27, p. 387, and $\Delta|I|^2_{mix}$ being the corresponding quantity for the mixer tube. Because the mixer-tube plate current is zero for half of each local-oscillator cycle, $\Delta|I|^2_{mix}$ is about half $\Delta|I|^2$. Also, g_c is approximately $1/\pi$ times g_m, and therefore R_{eq} for a mixer is approximately $(1/2)/(1/\pi)^2 \approx 5$ times the value of R_{eq} for the same tube employed as an amplifier.

The mixer of Fig. 56 is called a *grounded-cathode* or cathode-return mixer because the cathode serves as a common point for the input and output circuits. An alternate connection often convenient for use in high-frequency mixers is the *grounded-grid* or *grid-return* circuit, in which the r-f voltages are applied between cathode and grid and the i-f output voltage is obtained between plate and grid. A mixer circuit of this form is described in Art. 15, Chap. XII. The significant differences between the grounded-cathode and grounded-grid mixer circuits are as follows: (1) The grounded-grid circuit is easier to construct if the input frequency is so high that coaxial-line tuning elements are required. (2) The input impedance of the grounded-grid mixer is low because the signal source must supply the r-f component of plate current.

30. Automatic-frequency-control Circuits.[1] An automatic-frequency-control circuit, or *afc circuit*, is an arrangement for maintaining an essentially constant difference between the transmitter and local-oscillator fre-

[1] For a more detailed discussion of radar afc systems see R. V. POUND, *Microwave Mixers*, 290–351 (McGraw-Hill Book Company, Inc., New York, 1948), and S. N. VAN VOORHIS, *Microwave Receivers*, 27–78 (McGraw-Hill Book Compnay, Inc., New York, 1948).

quencies despite effects tending to shift these frequencies. The effects include temperature changes (especially important during first few minutes of operation) in both the transmitter and the local oscillator and transmitter load-impedance changes caused by antenna rotation. An afc circuit is seldom required in uhf systems but is often employed in 10-cm systems to counteract frequency shifts of the order of several megacycles per second and is nearly always used in 3- and 1-cm systems to avoid shifts of tens of megacycles per second.

The components of a typical afc circuit are indicated in the block diagram of Fig. 1*B*, p. 339. A small sample of the transmitted pulse is applied along with the local-oscillator voltage to a crystal mixer, and the mixer output after amplification by several i-f stages is applied to a special detector circuit called a *discriminator*. The d-c output of the discriminator is zero, positive, or negative, depending upon whether the intermediate frequency has the correct value or deviates from the correct value in one direction or the other. A larger direct voltage is produced by the control circuit and applied to the reflector electrode of the klystron local oscillator. The klystron frequency may be controlled within a small range by adjustment of the reflector voltage (see Art. 4, Chap. XI). Through the action of the discriminator and the control circuit, any deviation of the intermediate frequency from the desired value is made to bring about a reflector-voltage change that tends to restore the intermediate frequency to the correct value.

The afc mixer employs a crystal and is entirely similar to the signal-channel mixer. The afc i-f amplifier also is like the signal-channel i-f amplifier, except that the minimum possible noise figure is not required, and less gain is needed. The discriminator and control circuits are to be discussed in this article. In some afc systems, the afc attenuator, mixer, and i-f amplifier are omitted, and a voltage from the signal-channel i-f amplifier is applied to the discriminator. The receiver circuit is thereby simplified, but the frequency control obtained is less effective because the transmitted pulse in the signal channel contains extraneous frequency components introduced at the t-r device. Also strong signals from nearby radar equipment may appear in the signal channel with amplitudes comparable to the transmitted pulse and may influence the discriminator output.

A much-used discriminator circuit is drawn in Fig. 58. A coupling network connected to the final i-f amplifier tube V_0 supplies two i-f output voltages E_1 and E_2. Each of these voltages is applied to a diode detector (tubes V_1 and V_2), and the difference between the two direct voltages produced by the detectors is the output voltage E_{dc} of the discriminator. Inductors L_p and L_s are coupled by the mutual inductance M, and the very small capacitors C_1 and C_2 are used to tune the primary and second-

ary circuits. The coupling circuit is thus a special arrangement of the double-tuned circuit of Art. 23. Combinations of the primary voltage E_p with half the secondary voltage E_s form the i-f output voltages E_1 and E_2.

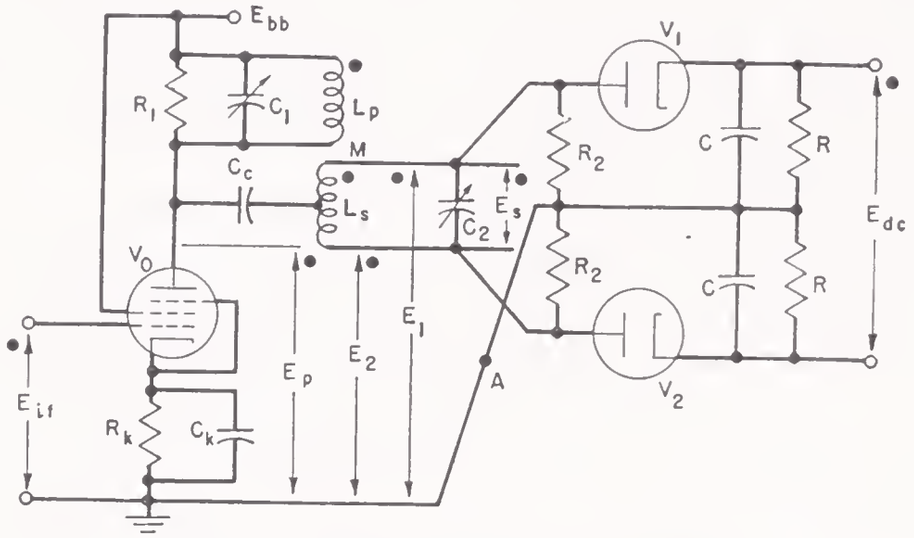
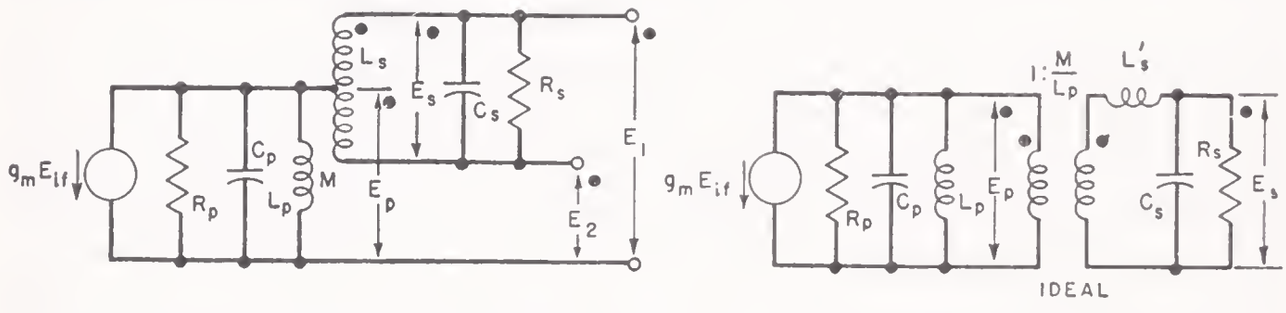


FIG. 58. Typical discriminator circuit.

The input impedances of the two diode detectors (see Art. 26) appear in series across the secondary circuit of the coupling network, and the sum of the two detector currents flows in C_c and is a load current for the primary circuit. Assume, temporarily, that the voltage drop in C_c is negligible and that the entire loading effect of the diodes can be represented by a load resistor across the secondary circuit. An equivalent circuit for the coupling network can then be drawn in the form of Fig.



(A) DIODE DETECTORS CONSIDERED EQUIVALENT TO SECONDARY LOAD RESISTANCE

(B) PRIMARY AND SECONDARY CIRCUITS SEPARATED

FIG. 59. Equivalent circuits for coupling networks in Fig. 58.

59*A*, in which R_p and C_p represent the total shunt resistance and capacitance of the primary circuit and R_s and C_s have similar significances for the secondary circuit.

A fact of particular interest is that at the desired intermediate frequency the secondary voltage E_s is 90 deg out of phase with the primary voltage E_p. This fact may be demonstrated by reference to the simplified equivalent circuit of Fig. 59*B*, in which the connection between primary and secondary circuits is omitted (because only the relation of E_s to E_p is to be considered) and the equivalent circuit of Fig. 46*A*, p. 129

in Chap. II, is employed for the coupling transformer. Inductance L_s' is equal to $(1 - k^2)L_s$, the short-circuit inductance of the secondary winding. A complex-number calculation based on Fig. 59*B* yields

$$\frac{E_s}{E_p} = \frac{M}{L_p} \frac{1}{1 - \omega^2 C_s L_s' + j\omega L_s'/R_s} \tag{68}$$

The primary and secondary tuning capacitors are adjusted so that with either winding short-circuited the other circuit is resonant at the desired intermediate frequency f_{if}. Thus $2\pi f_{if}$ corresponds to the mid-band angular frequency ω_0 employed in Art. 23. At the frequency f_{if}, L_s' and C_s are in resonance, $1 - \omega^2 C_s L_s'$ is zero, and by Eq. 68, E_s lags E_p in phase by 90 deg. At lower frequencies, $1 - \omega^2 C_s L_s'$ is positive and E_s and E_p differ in phase by less than 90 deg. At higher frequencies, the phase difference is greater than 90 deg.

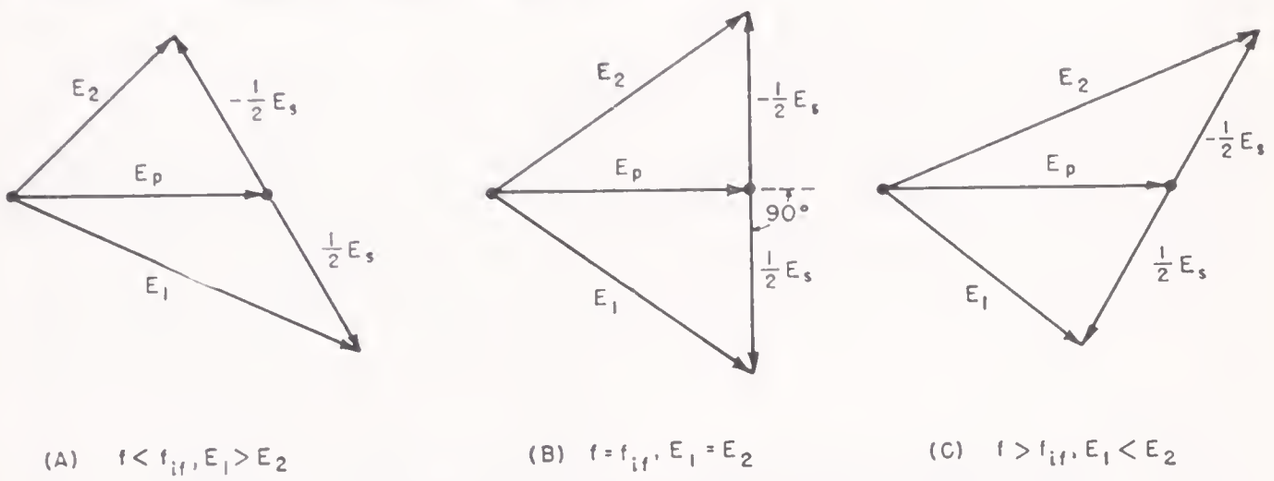
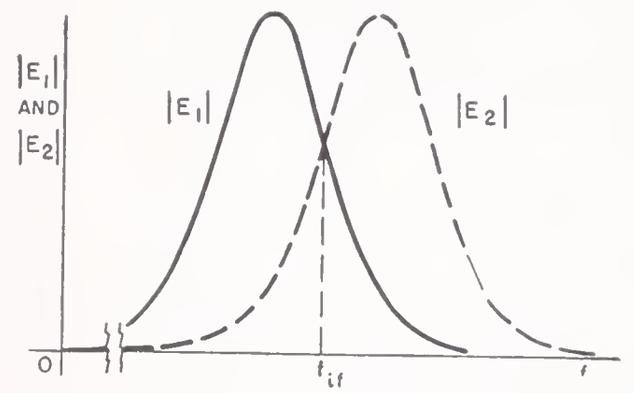


FIG. 60. Complex-number diagrams for network of Fig. 59*A*.

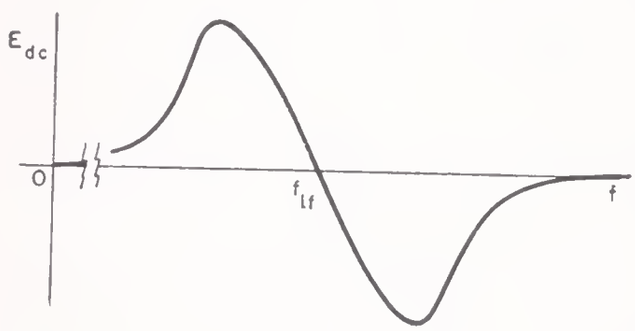
From Fig. 59*A* it may be seen that the i-f output voltages E_1 and E_2 equal, respectively, $E_p + \frac{1}{2}E_s$ and $E_p - \frac{1}{2}E_s$. Thus the complex-number diagrams of Fig. 60 apply, and the relative magnitudes of E_1 and E_2 depend in the manner indicated upon the frequency of the applied signal relative to f_{if}. The general way in which the magnitudes of E_1 and E_2 vary with frequency is indicated in Fig. 61*A*. The exact shapes of these curves depend upon the transformation ratio and coefficient of coupling of the transformer, upon the Q values for the primary and secondary circuits, and upon loading effects of the diode detectors not represented in the equivalent circuits of Fig. 59. Nevertheless, E_1 may be thought of qualitatively as the voltage developed across a parallel resonant circuit tuned to a frequency less than the desired intermediate frequency, and E_2 as the voltage across a similar circuit tuned to a frequency above the desired frequency.

The direct output voltages of the diode detectors in Fig. 58 are very nearly proportional to the i-f voltages E_1 and E_2, and therefore the shape of the curve of E_{dc} as a function of frequency may be determined by sub-

tracting the $|E_2|$ curve in Fig. 61*A* from the $|E_1|$ curve. The result, shown in Fig. 61*B*, is called the S curve of the discriminator and indicates that E_{dc} is zero for applied voltages of frequency f_{if}, and is positive for lower frequency voltages and negative for higher frequency voltages. Note that the frequency at which E_{dc} is zero is the frequency at which E_s is 90 deg out of phase with E_p and therefore depends upon the tuning of the discriminator secondary circuit. The symmetry of the two peaks of the S curve depends upon the discriminator-primary tuning or (if the S curve is drawn for the entire i-f amplifier combined with the discriminator) upon the tuning of all stages of the i-f amplifier.



(A) MAGNITUDE OF I-F OUTPUT VOLTAGE



(B) DISCRIMINATOR OUTPUT VOLTAGE

FIG. 61. Variation of discriminator voltages with frequency.

The waveform of the discriminator output voltage E_{dc} depends upon the time constant RC of the detector output circuits in Fig. 58. If the time constant is short compared with the period between pulses, the output consists of a series of video pulses, one for each transmitted pulse; if long, the output is an almost smooth direct voltage. The amplitude of the pulses, or the value of the direct voltage, varies with the frequency of the input to the discriminator in the manner indicated by the S curve of Fig. 61*B*.

Long-time-constant discriminator circuits are employed if the control circuit that connects the discriminator to the local oscillator is a d-c amplifier. A typical circuit of this type is indicated in Fig. 62. Note that, because a negative voltage of a few hundred volts must be supplied to the klystron reflector, the d-c amplifier is connected between ground and a negative supply voltage. The discriminator output circuit is at ground potential for i-f voltage components in Fig. 62, as in Fig. 58. The large negative direct voltage at the discriminator output does not affect the operation of the discriminator, because capacitor C_c is an open circuit for direct current, and the ground circuit of Fig. 58 is opened at point *A* when the connections of Fig. 62 are made.

The adjustable tap on resistor R_1 is set so that the voltage applied to the reflector of the local-oscillator tube when E_{dc} is zero results in an intermediate frequency of the desired value. This adjustment may be

made with the detector diodes V_1 and V_2 removed from their sockets. Often, in the OFF position of the afc ON-OFF switch, the voltage E_{dc} is short-circuited, and resistor R_1 is then used as the manual tuning control. Proper adjustment of the afc circuit is then secured by an initial manual

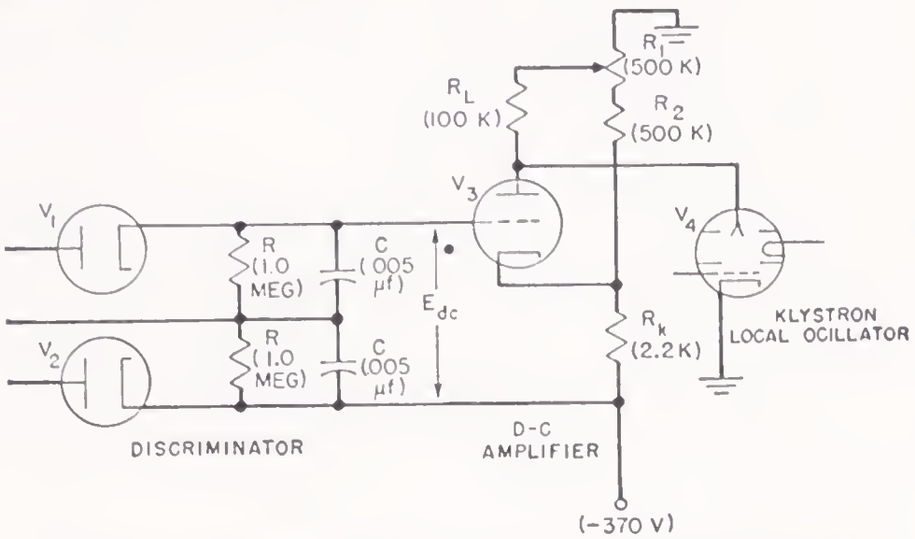
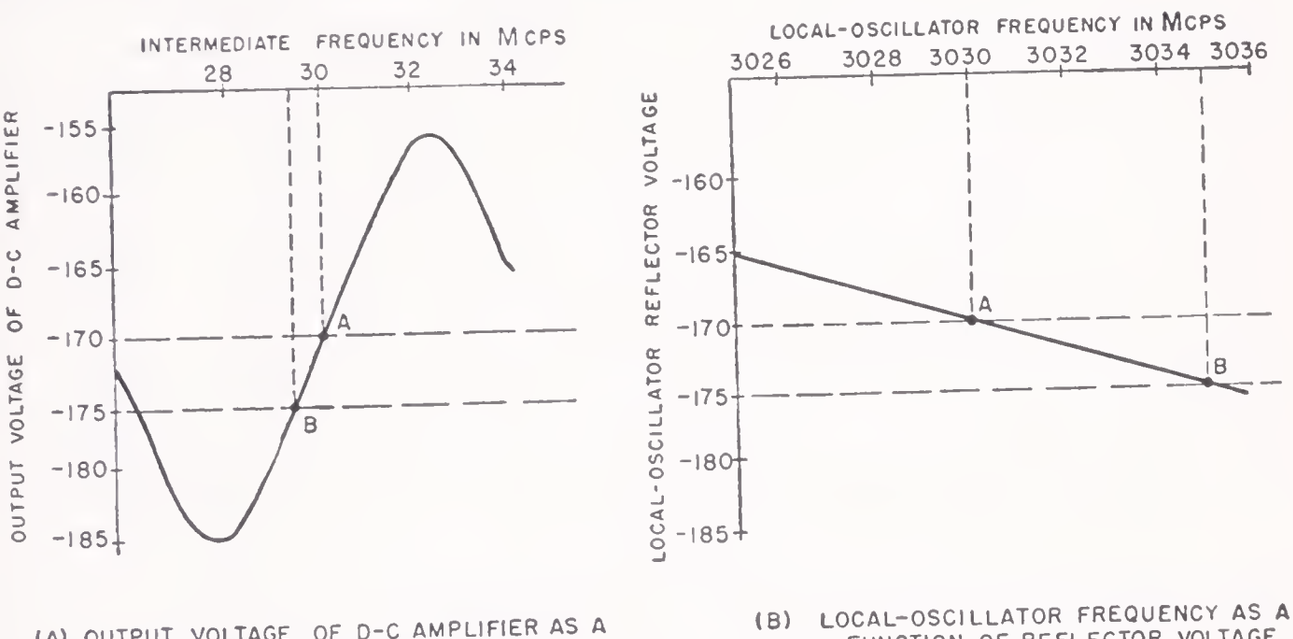


FIG. 62. Typical control circuit of d-c-amplifier type.

tuning of the receiver. When R_1 is correctly adjusted, the afc circuit has no effect upon the local-oscillator frequency unless the intermediate frequency changes from its correct value. If the intermediate frequency does change, the new voltage applied to the reflector of the local oscillator returns the intermediate frequency almost to its proper value.



(A) OUTPUT VOLTAGE OF D-C AMPLIFIER AS A FUNCTION OF INTERMEDIATE FREQUENCY

(B) LOCAL-OSCILLATOR FREQUENCY AS A FUNCTION OF REFLECTOR VOLTAGE

FIG. 63. Curves used in determining performance of afc circuit.

The frequency-stabilizing effect of the afc circuit may be determined from curves such as those of Fig. 63, which are typical for a radar system having a nominal transmitter frequency of 3,000 Mcps and a nominal local-oscillator frequency of 3,030 Mcps. The S curve of Fig. 63*A* is the amplified and inverted version of the discriminator S curve of Fig. 61*B*.

The variation of local-oscillator frequency with reflector voltage is indicated in Fig. 63*B*, and it is assumed that the klystron oscillates strongly over the reflector-voltage range of −150 to −190 volts. (See Art. 4, Chap. XI, for a discussion of reflector-voltage tuning of klystrons.) For a transmitter frequency of 3,000 Mcps, operation is at point *A*, the local-oscillator frequency being 3,030 Mcps, the intermediate frequency 30 Mcps, and the reflector voltage −170 volts.

If the transmitter frequency increases, the intermediate frequency decreases and the action of the discriminator makes the repeller voltage shift in a negative direction. This shift increases the local-oscillator frequency and almost compensates for the increase in transmitter frequency. Suppose, for example, that the transmitter frequency shifts by an amount which decreases the intermediate frequency by 0.5 Mcps. Operation is then at point *B* in Fig. 63*A*, and the local-oscillator reflector voltage is −175 volts. This reflector voltage, according to the curve of Fig. 63*B*, produces a local-oscillator frequency of 3,035 Mcps. Since the transmitter frequency is assumed to be lower than that of the local oscillator, it is 3,035 − 29.5 = 3,005.5 Mcps; that is, a change of 5.5 Mcps in transmitter frequency is required to shift the intermediate frequency by 0.5 Mcps.

A disadvantage of the circuit of Fig. 62 is its relatively slow operation. Very sudden changes of transmitter frequency may occur because of impedance changes caused by antenna rotation. The circuit of Fig. 62 does not correct for these changes immediately because of the long time constant employed in the discriminator output circuit. Another disadvantage, which is of considerable importance in 3- and 1-cm systems, is the need for an initial manual-tuning adjustment before automatic operation can begin. Control circuits employing gas triodes have been devised to provide automatic lock-in of the afc circuit and also somewhat more rapid response to frequency changes.

One form of gas-tube control circuit is indicated in simplified form by the diagram of Fig. 64. The discriminator time constant is short, so that positive or negative video pulses are obtained from the discriminator, and these pulses are amplified and inverted by the video amplifier of Fig. 64. The discriminator connections for Fig. 64 are reversed relative to those of Fig. 62; thus the video-amplifier output pulses are positive for intermediate frequencies less than the correct value and negative for higher intermediate frequencies. These pulses are applied to the grid of gas triode V_3 in Fig. 64.

To understand the operation of this circuit, first assume that the radar transmitter is off. When power is applied to the afc circuit in Fig. 64, zero initial voltages exist across capacitors C_1, C_2, and C_3, and tube V_4 conducts because its grid-to-cathode voltage is zero. As capacitor C_2

charges through the path shown in Fig. 64, the plate-to-cathode voltage of V_4 decreases until a value is reached which is insufficient to sustain conduction. This occurs when the voltage across C_2 reaches -200 volts, as indicated in Fig. 65. As soon as tube V_4 stops conducting, capacitor C_2 discharges through resistors R_4 and R_5 until the plate-to-cathode voltage rises sufficiently to permit conduction. The voltage at which V_4

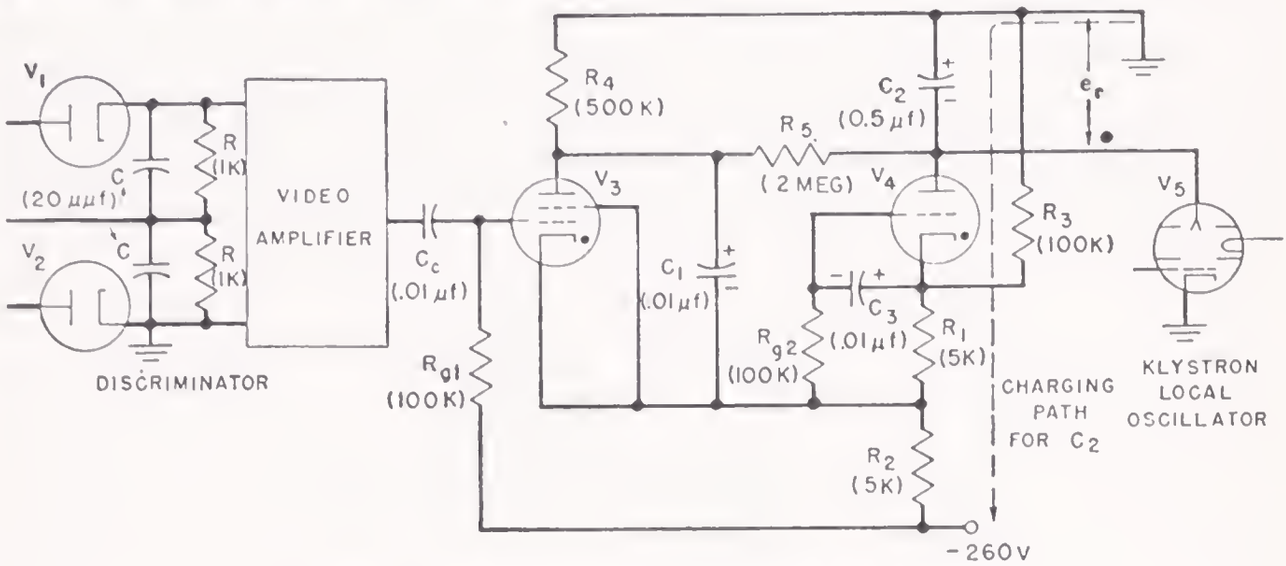


FIG. 64. Control circuit employing gas triodes.

conducts depends upon the magnitude of the negative grid-to-cathode voltage built up across C_3. For the circuit values in the figure, the voltage e_r across C_2 reaches -100 volts when V_4 begins to conduct. Thus a saw-tooth waveform having an amplitude of 100 volts and a frequency of about 1 cps (see Fig. 65) is applied to the reflector of the local oscillator. The negative grid-to-cathode voltage developed across R_2 prevents V_3 from conducting when the radar transmitter is off.

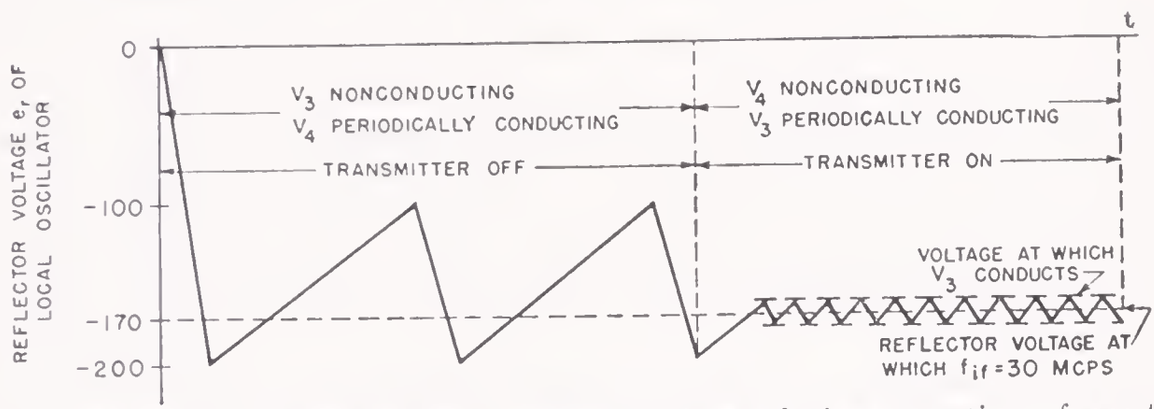


FIG. 65. Variation of local-oscillator reflector voltage during operation of gas-triode circuit.

When the reflector voltage is periodically varied, the local-oscillator tube oscillates during only a portion of the sweep. Assume that oscillations occur in a range of about -150 to -190 volts and that the curve of Fig. 63*B* applies. When the transmitter is turned on, an intermediate-frequency signal is produced and, for a transmitter frequency of 3,000 Mcps, varies from a value somewhat above 30 Mcps to one somewhat

below 30 Mcps as the reflector voltage rises from −190 to −150 volts. Because of the discriminator circuit, video pulses appear at the video-amplifier output in Fig. 64 when the reflector voltage is near −170 volts, and the variation of the pulse amplitudes is as indicated in Fig. 66.

When the voltage on the reflector of the local oscillator reaches a value slightly above −170 volts, the intermediate frequency is somewhat less than 30 Mcps and the pulses applied to the grid of V_3 in Fig. 64 are positive and have sufficient amplitude to cause the tube to conduct. As a result, capacitor C_1 discharges quickly through the tube. Tube V_3 stops conducting when its plate-to-cathode voltage reaches about 20 volts, and C_1 begins to recharge through resistor R_4.

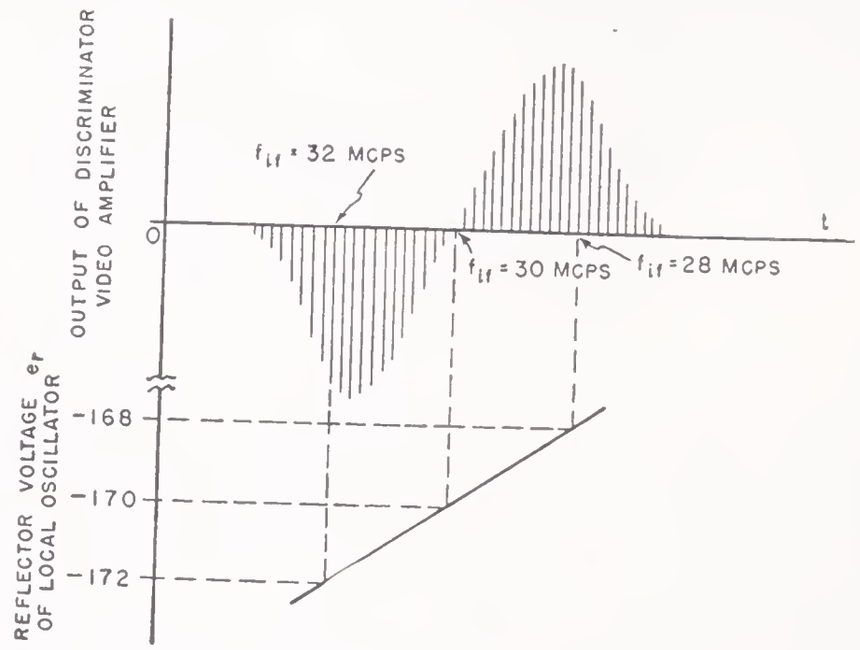


FIG. 66. Variation of video-pulse amplitude in gas-triode circuit.

The abrupt drop in voltage at the plate of V_3 causes C_2 to begin to charge through R_5 and C_1, thus interrupting the exponential rise of the reflector voltage e_r and starting a decrease of e_r. The decreased value of e_r increases the intermediate frequency and removes the positive pulses from the grid of V_3. Therefore, C_1 is able to recharge through R_4, and the rise of the reflector voltage begins again. When the intermediate frequency again becomes less than 30 Mcps, V_3 conducts once more and the cycle repeats. The reflector voltage therefore varies in a small range around −170 volts. It has an approximately saw-tooth waveform, a repetition rate of about 200 cps, and an amplitude of only a fraction of a volt. It is illustrated by the small-amplitude voltage variations in Fig. 65. Because this voltage is applied to the reflector of the local-oscillator tube, the output frequency of the receiver mixer is not constant. However, the variations lie in a range centered at 30 Mcps and of width less than the i-f-amplifier bandwidth.

While the afc circuit is locked in and V_3 is conducting periodically, V_4 does not conduct and plays no part in the operation of the circuit. This

tube functions only to bring the reflector voltage into the correct range and to reestablish correct operation if the afc circuit is thrown out of the correct range at any time during operation—by an abrupt change of transmitter frequency, for example.

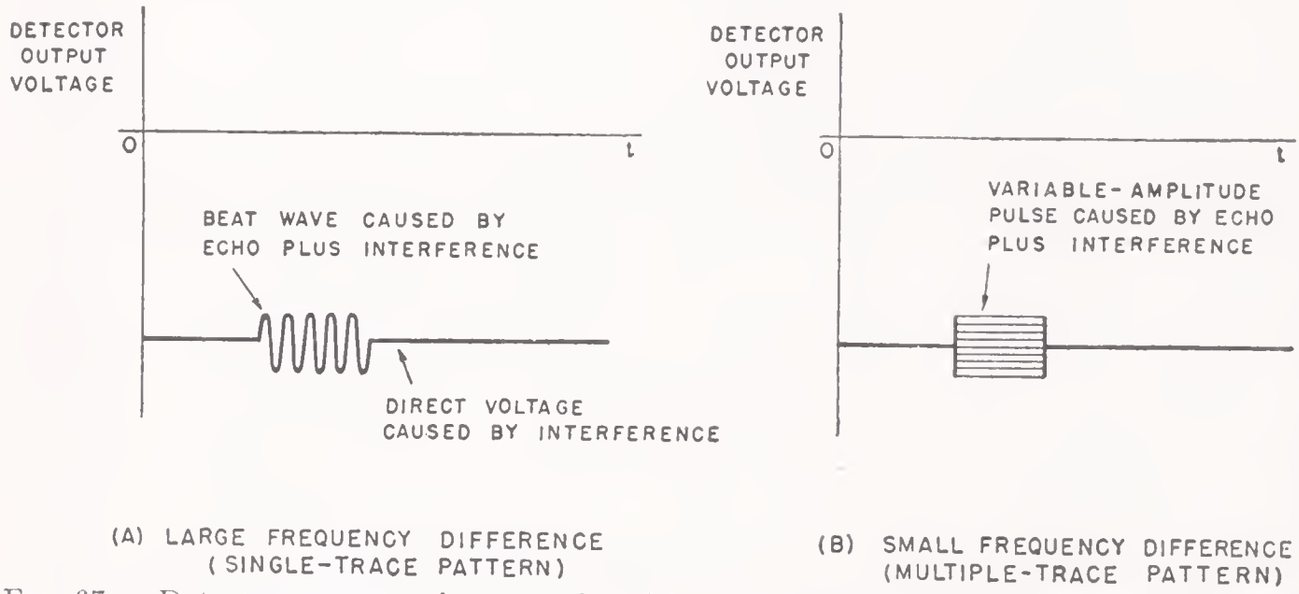
31. Anti-interference Circuits and Signal Discernibility. As explained in Art. 17, the minimum amplitude of a discernible signal depends upon (1) the amount of noise and interference received by the antenna, (2) the amount of noise generated in the receiver input circuits, and (3) the effects of receiver circuits and indicator presentation upon signal discernibility. Input-circuit noise generation (as measured by noise figure) is discussed in Arts. 17 and 25, and the effects of bandwidth and detector nonlinearity upon signal discernibility are considered in Arts. 18 and 26. The present article treats special circuits sometimes added to a receiver to improve the discernibility of signals in the presence of strong interference.

Echo-pulse signals may be obscured not only by fluctuation noise received by the antenna or generated within the receiver but also by a variety of interfering signals arising from near-by radar systems or other electrical equipment. Such interfering signals frequently are of much greater amplitude than the noise, but fortunately they have waveforms of such a nature that it is often possible to reduce considerably the extent to which echo pulses are obscured. The interfering signals may take the form of modulated or unmodulated sine waves of frequencies near that of the radar system (other frequencies are rejected by the receiver) or of transient disturbances having frequency components in the range accepted by the receiver. Another form of interference known as *clutter* arises if the echo pulse from an object of interest is obscured by signals returned from other objects. Sea return, for example, is a form of clutter and may interfere with the detection of echoes from ships or from aircraft flying over water. Clutter received from land masses may interfere similarly with signals from aircraft over land, and even clouds and rainstorms may cause clutter that increases the difficulty of aircraft detection, particularly at radar wavelengths of 3 cm or less.

Consider first that the interfering signal is an unmodulated sine wave, and assume that the interfering signal and the echo pulse of interest have different frequencies. The interfering signal alone produces a direct voltage at the detector output, and the beat effect that occurs when interference and echo pulse are present simultaneously results in a detector output wave of the general form illustrated in Fig. 67. (A detector that normally produces negative pulses is assumed.) The form indicated in Fig. 67*A* occurs if the frequency difference between interference and echo pulse is several megacycles. If the frequency difference is very small, the phase difference between interference and echo changes

little during the pulse duration but has different values on successive range sweeps, so that the pattern of Fig. 67*B* is obtained. These patterns show that some indication of the presence of the echo pulse is obtained even though the interfering signal has an amplitude many times that of the desired signal.

The patterns of Fig. 67 are obtained only if the i-f amplifier is not saturated by the interfering signal. If any stage saturates, its output is unaltered by the echo pulse, and no indication of the desired signal appears in the receiver output. The primary requirement for detecting small signals in the presence of interference is as follows: *The gain of the i-f amplifier must be reduced sufficiently to prevent the interfering signal from saturating any stage of the receiver.* Many anti-interference circuits are



(A) LARGE FREQUENCY DIFFERENCE (SINGLE-TRACE PATTERN)

(B) SMALL FREQUENCY DIFFERENCE (MULTIPLE-TRACE PATTERN)

FIG. 67. Detector output voltage produced by combination of unmodulated interfering signal with echo-pulse signal.

merely special means of reducing the i-f gain at the proper times and by appropriate amounts.

Because gain must be decreased to avoid saturation, the echo pulses that can be detected in the presence of an unmodulated interfering signal, though considerably smaller than the interfering signal, are, of course, much larger than those which could be detected in the absence of interference. If the grid-bias form of gain control (see Art. 22) is employed, the nonlinear characteristic of the tubes can be used to reduce the interfering signal greatly by cutoff clipping and at the same time to provide relatively high gain to the changes in the envelope of the interference wave produced by the addition of the echo pulse.

Even though saturation of the i-f amplifier is avoided, signal indications may still be lost if no coupling capacitor is employed between the detector and the video amplifier—the first video stage may be biased beyond cutoff by the d-c component of the detector output (see Fig. 67). To avoid this possibility, most radar sets employ a coupling capacitor.

(The capacitor cannot cause blocking if the detector output voltage is negative.) If the frequency difference between the interference and the signal is large, the frequency of the beat wave in Fig. 67*A* is high and a very wide-band video amplifier is required if signals are to reach the indicator. Sometimes arrangements are made to switch an extra detector into the circuit when interference is present so that the beat-frequency pulse may be rectified and a lower frequency signal applied to the video amplifier and indicator.

If the interfering signal is modulated at a low frequency or occurs in pulses of long duration, the same considerations apply as for an unmodulated interference wave, except that for optimum echo detection the gain should be increased and decreased in accordance with the amplitude variations of the interference wave. Such rapid variation of gain is not possible by manual control but may be obtained with the aid of special circuits.

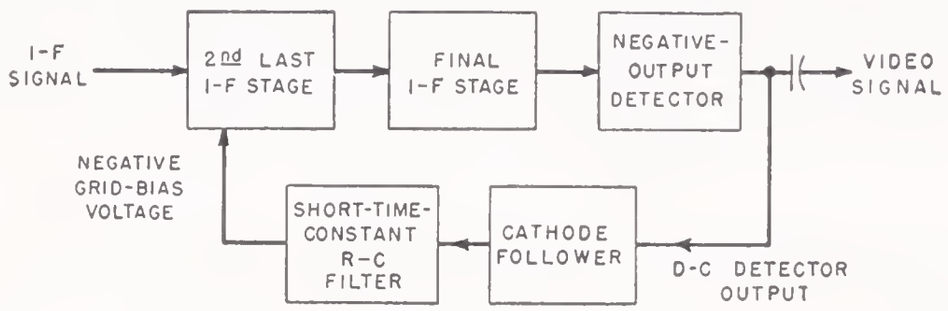
One form of gain control that may be provided by these special circuits is known as *automatic back bias*, or *instantaneous automatic gain control*. Typical circuits are illustrated in block-diagram form in Figs. 68*A* and *B*. In each circuit a negative direct voltage that varies in magnitude in accordance with the amplitude of the interfering signal is produced by a detector and is used to control the bias voltage of an i-f stage, thus automatically reducing the gain with increase of the amplitude of the interfering signal.

The circuits of Fig. 68 differ from the conventional automatic-gain-control circuits employed in broadcast receivers (and in some special-purpose radar receivers) in that the time constant of the *R-C* filter through which bias voltage is applied to the i-f amplifier is short—of the order of a few tens of microseconds—rather than long enough to remove even low audio frequencies from the gain-control voltage. The time constant is short enough to permit the gain-control voltage to follow low-frequency modulations of the interference, yet long enough to prevent the gain-control voltage from changing during reception of an echo pulse of a few microseconds duration. Because the filter time constant must be short, the difficulty of preventing oscillations in the feedback loop formed by the gain-control circuit is considerable. Ordinarily it is possible to include at most two stages in the feedback loop (as in Fig. 67*A*) or one stage if a video amplifier is used in the control circuit (as in Fig. 67*B*).

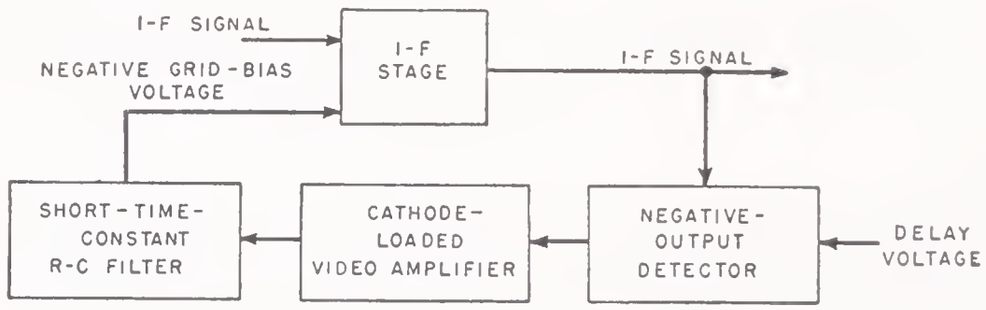
The circuit of Fig. 68*A* makes use of the signal detector of the receiver in the gain-control circuit and applies bias voltage to the grid of the next to last stage of the i-f amplifier. Thus only the final two i-f stages are protected from saturation. The circuit of Fig. 68*B* is somewhat more complicated because a separate detector for the gain-control circuit is used, but it can be connected to any desired i-f stage. Often several

circuits like the one in Fig. 68*B* are used in a receiver, one being connected to the final i-f stage and others to as many earlier stages as require protection. When the separate detector of Fig. 68*B* is employed, a delay voltage may be applied to the detector diode. The delay voltage makes the plate of the diode negative with respect to the cathode, so that small i-f signals cause no detector output. Thus full i-f gain is maintained except for signals that approach the saturation level.

When the interfering signal is modulated, special circuits are required not only to avoid i-f amplifier saturation but also to avoid video-amplifier saturation. The "d-c" output of the detector (see Fig. 67) changes in



(A) CONTROL ACROSS TWO STAGES



(B) CONTROL ACROSS ONE STAGE

FIG. 68. Block diagrams of typical circuits for instantaneous automatic gain control.

magnitude in correspondence with the modulation envelope of the interfering wave and therefore causes variations of the bias voltage for the first video-amplifier stage, despite the use of a detector-output coupling capacitor. In consequence, the first video stage may be biased beyond cutoff during a portion of the modulation cycle. To avoid this effect some receivers are provided with a short-time-constant coupling circuit at the detector output. The time constant is made a few tens of microseconds, so that brief echo pulses are not altered by the coupling circuit, but low modulation frequencies are greatly attenuated.

Both the automatic-back-bias circuit and the short-time-constant coupling circuit produce the effect of *edge response* in the receiver. That is, on a plan-position indicator, for example, small objects are presented in the normal manner but large objects are marked only by a line along the edge of the object nearest the radar set. Such a display is more

difficult to interpret than a normal indicator presentation. Therefore switching arrangements are employed so that the back-bias circuit can be disconnected and the coupling-circuit time constant increased when interference is not present.

Another scheme for automatic control of i-f-amplifier gain, called *sensitivity-time control* because the gain is varied as a prescribed function of time, is useful when the interference is the clutter produced by land or sea return. Clutter-type interference takes the form of relatively long pulses, one pulse being produced for each range sweep. Each pulse has high amplitude shortly after the time of the transmitted pulse (because of the great intensity of the return from near-by objects) and decreases in amplitude more or less as the square of the time interval corresponding to range, finally becoming negligible after an interval corresponding to a range of a few tens of miles. For maximum discernment of signals amidst clutter, therefore, the receiver gain should be low at the beginning of each sweep and should increase continuously as the sweep progresses. To provide the required variation of gain, a sensitivity-time-control circuit takes the form of an R-C circuit somewhat similar to a saw-tooth generator (Art. 5, Chap. II) synchronized by the transmitted pulse. This circuit supplies a large negative bias voltage to one or more i-f stages at the beginning of each range sweep, and the magnitude of the bias voltage decreases because of the discharge of a capacitor as the sweep progresses.

A disadvantage of the sensitivity-time-control circuit is that adjustment of the circuit constants for optimum variation of gain with time under various clutter conditions is difficult. The automatic-back-bias circuit provides the possibility of a more precise control of gain, but it does not operate well when required to produce the very large gain changes needed when clutter is present. Particularly effective is a combination of the two gain-control circuits in which the sensitivity-time-control circuit is adjusted to provide somewhat more than optimum gain throughout the range sweep. Errors in the sensitivity-time control then become unimportant because the back-bias circuit provides whatever additional gain reduction is needed at each instant during the sweep.

To prevent i-f-amplifier saturation, a procedure alternate to that of adding special gain-control circuits to the amplifier is the use of a nonsaturating amplifier often called a *linear-logarithmic amplifier*. This amplifier provides an output voltage that is a linear function of the input voltage for low-amplitude signals and is approximately a logarithmic function for high-amplitude signals. That is, the linear range does not terminate in a definite saturation point; rather, continued increase of the input signal causes continued increase in the output, though at a much reduced rate. Application of a large interfering signal to a linear-

logarithmic amplifier therefore does not fully saturate the circuit but merely reduces the extent to which the circuit amplifies a simultaneously applied, small echo pulse. The net result is the same as could be obtained in an ordinary amplifier by decreasing the amplifier gain to avoid saturation. Because the effective gain for the small echo pulse is the slope of the output-input curve and the derivative of $\ln x$ is $1/x$, the logarithmic output-input curve provides a signal gain inversely proportional to the amplitude of the interference.

One circuit for approximating the linear-logarithmic characteristic is illustrated in block-diagram form in Fig. 69. The i-f stages are so constructed that the amplitude of the input voltage remains essentially constant at the saturation value when the input-signal amplitude is made greater than that which causes saturation. If detectors 2 and 3 were

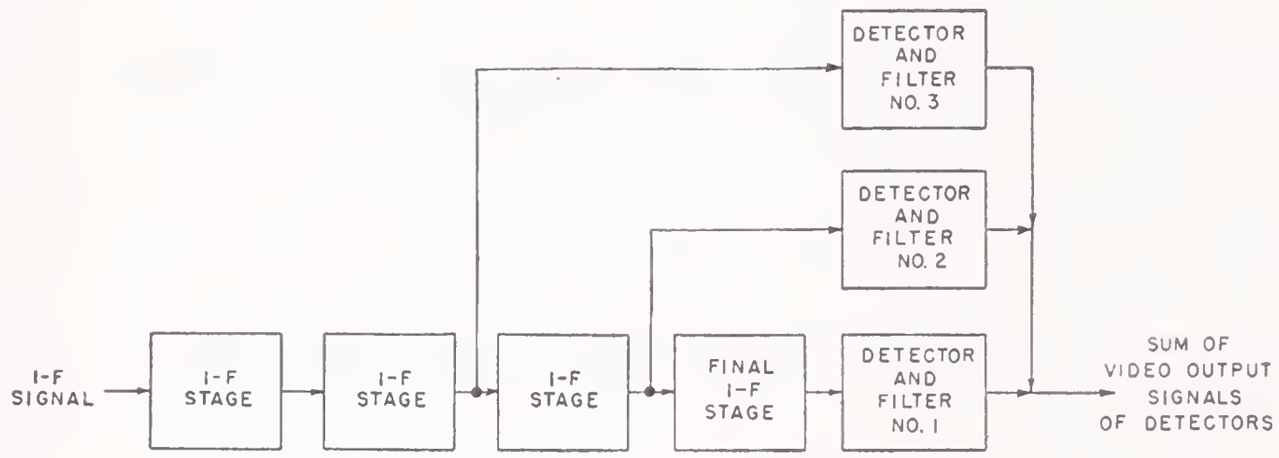


FIG. 69. Block diagram of a linear-logarithmic amplifier.

not present, the output voltage would be limited by saturation of the final i-f amplifier. Larger signals would cause an increased output from the next to last stage, however, and in Fig. 69 this increase is made to appear at the output by the addition of the detector 2 signal to the output of detector 1. Similarly, use of detector 3 causes the output to continue to increase after the next to last stage saturates. The gain becomes less as each stage saturates, and by proper choice of stage gains and points of saturation an approximately logarithmic over-all curve can be had. The number of parallel signal channels required depends upon the range of amplitudes over which the logarithmic characteristic is required to extend.

The linear-logarithmic amplitude characteristic is particularly desirable when clutter interferes with the discernment of echo pulses. Clutter may be considered to comprise an erratic variation wave not unlike fluctuation noise superposed on a pulse that begins at high amplitude and decreases in amplitude with increase of range. The fluctuating component obscures the signal in the same way as would a noise wave of great amplitude and thus limits the receiver amplification that may usefully be employed. The clutter fluctuation is roughly proportional to

the total clutter amplitude, and therefore for constant fluctuation amplitude at the receiver output, the amplification should be inversely proportional to the clutter amplitude. Exactly this relation is provided by a logarithmic output-input curve.

The random variations of fluctuation noise and clutter fluctuations are especially undesirable because of the extreme difficulty of detecting signal in the presence of such waveforms. However, if observations of signals received from a particular direction can be made over an extended time interval, the discernibility of signals amid noise or fluctuation-type interference can be greatly increased by an averaging process often called *signal integration*. The waveforms of received signals corresponding to many successive range sweeps are added point by point along the range trace. Any echo-pulse signals that may be present appear in each trace and are built up to many times the single-trace amplitude by the addition. The fluctuation waveform, on the other hand, is random and does not repeat from trace to trace. Thus the fluctuation amplitude is not increased nearly so much as the pulse-signal amplitude in the addition process, and the amplitude of a discernible signal becomes smaller as the number of traces that are combined is increased. The increased discernibility is obtained at the cost of a longer period of observation.

Circuits for producing an addition of this kind can be made, but a simpler procedure, usually, is to perform the additions at the indicator screen. A plan-position indicator, for example, having a long-persistence screen and used with a very slowly rotating antenna develops at each point a brightness that depends upon the combined effects of many nearly identical range traces. Even when a Type A indicator with a short-persistence screen is employed, a considerable integrating effect is obtained in the eye and mind of the observer. Devices to increase signal discernibility by integration must add signals over a much longer time than does the eye if an appreciable improvement is to be obtained. A result of the signal-integration process is that the speed of rotation of the antenna, the antenna beamwidth, and the repetition rate of the set all have important effects upon signal discernibility.

CHAPTER VII

RADIO-FREQUENCY TRANSMISSION LINES

The radar components discussed in the preceding chapters operate at frequencies below a few tens of megacycles per second and are treated for the most part as *lumped-parameter circuits*, that is, as networks of vacuum tubes and resistance, capacitance, and inductance elements. The components that will be described in the remaining chapters—transmission lines, waveguides, resonators, and antennas—are *distributed-parameter devices*. The frequencies of operation are so high that the inductances of short lengths of conductor and the capacitances between short conductors and their surroundings cannot be neglected. These inductances and capacitances are distributed along the length of a conductor, and their effects intermingle at each point of the conductor. At frequencies so high that wavelength is comparable to the physical dimensions of apparatus, distributed parameters cannot be represented accurately by means of lumped-parameter equivalent circuits. Frequently, in such circumstances, an analysis that considers directly the *electric* and *magnetic fields* in the device is preferable to one based upon circuit concepts. Transmission lines form a convenient link between low-frequency circuit concepts and high-frequency field concepts, because they can be analyzed either in terms of voltage and current or in terms of electric and magnetic fields.

The high-frequency radar components considered in the remaining chapters are discussed in terms of sinusoidal variations with time of voltage and current or of electric and magnetic fields. Normally, the high-frequency sine waves in radar sets are modulated with an approximately rectangular envelope. Sine-wave analysis of linear high-frequency devices is adequate even though modulated sine waves are applied, because, as explained in Art. 4 of Chap. VI, such waves may be divided into sinusoidal components of different frequencies and amplitudes. The frequencies of the several components lie in a band centered about the carrier frequency, the width of the band being only a very small fraction of the center frequency.

The r-f transmission system of a radar set may employ either two-conductor lines or waveguides. Two-conductor lines are used in uhf sets and in some 10-cm sets, whereas waveguides are sometimes employed at 10 cm and always at 3 and 1 cm. This chapter is limited to a discussion

of two-conductor lines; the subject of waveguides is presented in Chap. VIII.

1. Types of Transmission Lines. The earliest radar systems operated at frequencies of the order of 100 Mcps and employed transmission lines consisting of two parallel tubular conductors of copper or brass supported at intervals by means of stand-off insulators. The value of the characteristic resistance (Art. 2) of such lines ranges from 100 to 300 ohms. A typical 200-ohm line, for example, consists of two $\frac{3}{8}$-in. conductors spaced 2 in. between centers. Parallel-wire lines are *balanced-to-ground* devices; that is, in normal operation a positive voltage (relative to ground or a shielding enclosure for the line) at a point of one conductor is balanced by an equal-magnitude negative voltage at the corresponding point of the other conductor. These lines are therefore most conveniently used with other balanced-to-ground devices such as push-pull vacuum-tube circuits. Although a parallel-wire line is simple to build, its use in radar is limited to sets operating below about 200 Mcps. Above that frequency, power loss due to induction and radiation becomes excessive.

A more suitable line at higher frequencies is the coaxial line, composed of two concentric conductors separated by an insulating material. Because the electromagnetic field associated with a coaxial line is confined to the space between inner and outer conductors, neither induction nor radiation loss occurs. Coaxial lines have low attenuation even at very high frequencies and are used extensively at 3,000 Mcps and all lower frequencies. Their characteristic resistance is much lower than that of parallel-wire lines, being of the order of 50 to 70 ohms. Unlike a parallel-wire line, the coaxial line is unbalanced to ground and may therefore be connected directly to single-tube circuits.

Coaxial lines for use in radar are constructed in both flexible and rigid forms. The inner conductor of a flexible cable is surrounded throughout its length by a flexible plastic material (polyethelene, for example). A braid of conducting material woven over the dielectric serves as the outer conductor. A protective covering such as vinylite is usually placed over the conducting braid, and sometimes the entire cable is covered with a steel braid or armor. The armor provides additional mechanical strength without affecting the electrical characteristics of the cable. Constructional details of a coaxial cable are shown in Fig. 1*A*, and physical and electrical characteristics of commonly used cables are given in Table 1, p. 544.

At frequencies near 3,000 Mcps, flexible coaxial cables have the disadvantage that losses in the dielectric material become appreciable, especially in installations that require long cables. Many high-frequency systems therefore employ rigid coaxial lines with air as the dielectric.

The rigid line consists of a tubular inner conductor supported concentrically within a large outer conductor by means of either *beads* or *quarter-wave stubs.*

A portion of a bead-supported line is drawn in Fig. 1*B*. For operation in the uhf range, the beads are frequently ceramic, but in the microwave region they are more commonly made of polystyrene in order to reduce the dielectric loss in the beads. In the use of a bead-supported line, care must be taken to avoid insulator breakdown because of excessively high voltages that may be present in the line. (Even in low-power systems, very high voltages may occur if adjustments at the line terminations are incorrect.) Breakdown usually occurs at the beads because the intensity

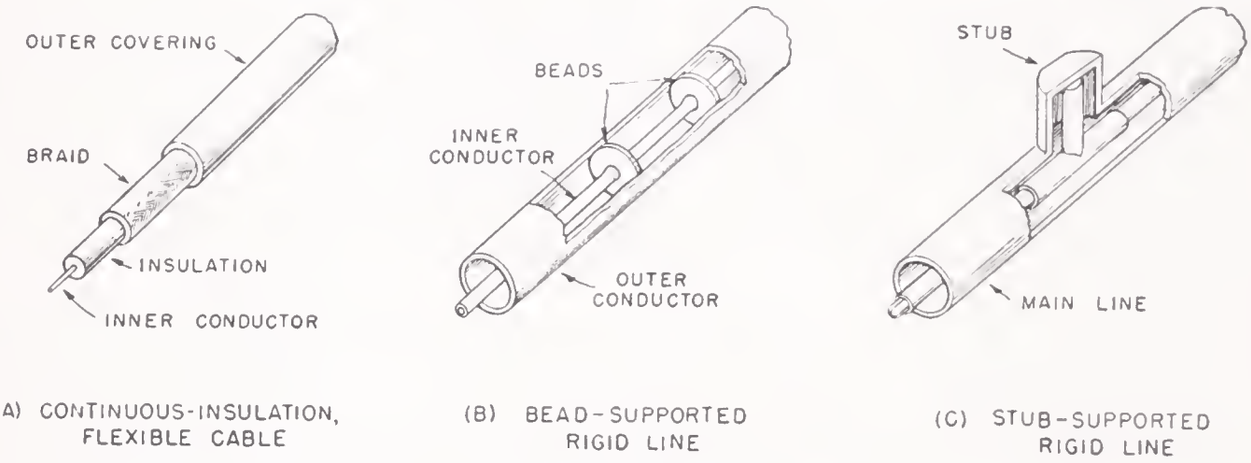


FIG. 1. Types of coaxial lines.

of the electric field is higher in the minute air spaces between beads and conductors than it is elsewhere. Dust and moisture on the surface of the beads may also provide a leakage path and thus cause breakdown. Polystyrene beads have the disadvantage that carbonization of the polystyrene by the arc occurring at breakdown is likely to cause a permanent fault and necessitate replacement of the bead. Another disadvantage of beads, particularly at microwave frequencies, is that they cause appreciable reflection of the electromagnetic waves within the line. It is possible, however, to minimize the reflections by use of thin beads and by the use of undercut beads[1] or a strategic spacing of the beads along the line.[2]

A section of stub-supported line is indicated in Fig. 1*C*. The stubs are short sections of shorted-end transmission lines in parallel with the main line. The stub length is chosen so that an extremely high impedance is presented to the main line at the operating frequency (see Art. 3). The presence of the stubs therefore has negligible effect upon the performance of the main line. Stub-supported lines have very low attenuation

[1] If the diameter of the inner conductor is decreased at the bead, the bead is said to be undercut. See R. W. CARNES, "A Coaxial-line Support for 0 to 4000 Mcps," *Proc. I.R.E.*, **37** (January, 1949), 94–97.

[2] G. L. RAGAN, *Microwave Transmission Circuits*, 155–170 (McGraw-Hill Book Company, Inc., New York, 1948).

because of the absence of dielectric material. Their use is generally limited to microwave frequencies, however, because at lower frequencies the lines become unwieldy owing to the excessive stub lengths required. Further details of stub supports are discussed in Art. 5 of Chap. XII.

2. Traveling Waves. In Art. 22 of Chap. II, it is shown that, when a battery of voltage E is suddenly connected to an uncharged lossless transmission line, the voltage between conductors does not jump instantly to a voltage E at all parts of the line. Instead, the input-terminal voltage rises instantly to E, and a steep voltage wavefront of magnitude E travels from the battery toward the far end of the line. If the inductance and capacitance distributed along a unit length of line are represented by l and c, respectively, the wavefront travels at a speed

$$v = \frac{1}{\sqrt{lc}} \tag{1}$$

and reaches a point remote from the battery after a time interval proportional to the distance between the battery and the point. For usual line lengths, the time interval is exceedingly short because, as explained in Chap. II, the speed of propagation is of the same order of magnitude as the speed of light.

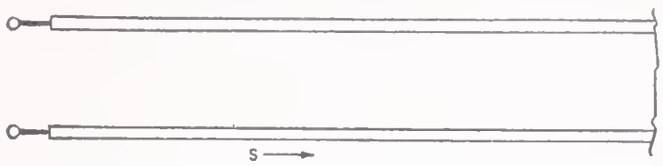
The voltage wavefront is accompanied by a similar wavefront of current. Both travel in step, and at any point of the line, the voltage between conductors at any instant is a constant times the current in the conductors. This constant, called the characteristic resistance R_c, is a property of the transmission line and is given by the relation

$$R_c = \sqrt{\frac{l}{c}} \tag{2}$$

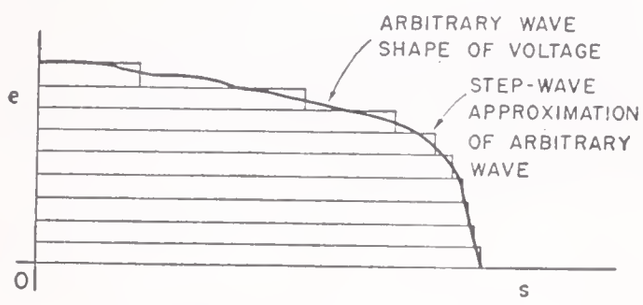
In the r-f components of radar systems, sine-wave generators rather than batteries are connected to transmission lines and sine waves instead of step waves travel along the lines. A fact of fundamental importance is that a wave of *any shape* is propagated without change of shape or magnitude along a uniform lossless line at a speed given by Eq. 1. The voltage wave is always accompanied by a current wave of similar shape, and at a given instant the ratio R_c of the amplitudes of the two waves at any point has the value indicated by Eq. 2. The behavior of uniform lossless lines is the subject of most of this chapter; the effects of small losses are considered in Arts. 13 and 14.

That any form of wave can be propagated by a transmission line may be shown in one way by representing an arbitrary wave shape, such as the voltage wave of Fig. 2, as the sum of many properly located step waves of small amplitude. Because the line is a linear system, each step wave of

voltage may be considered separately. All component step waves travel with the same velocity (specified by Eq. 1) and maintain constant amplitude. The component waves therefore travel as a unit, and their sum is a voltage wave that maintains constant shape and amplitude as it moves along. Thus the wave that the small step voltages approximate is also propagated without change of shape or amplitude.



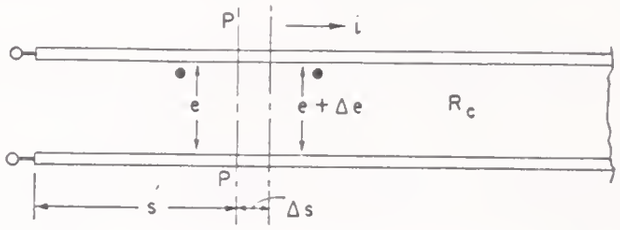
(A) SECTION OF TRANSMISSION LINE



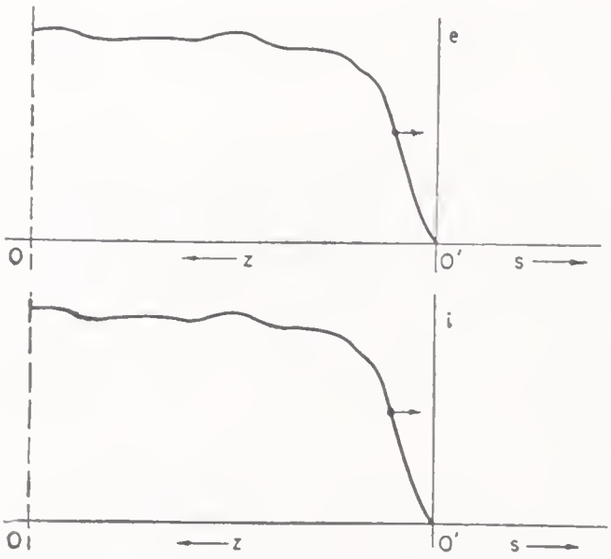
(B) TRAVELING WAVE OF VOLTAGE

FIG. 2. Propagation of a wave of arbitrary shape on a lossless line.

That the velocity of propagation is $1/\sqrt{lc}$ and the voltage-current ratio is $\sqrt{l/c}$ for waves of any shape may be shown in another way with the aid of the diagram of Fig. 3. Voltage and current waves in this diagram are assumed to move away from the input terminals of the line with constant velocity and without change of shape or amplitude. The velocity of the waves is called v, the distance of a point P from the input terminals is s, and the voltage and current at P for any instant of time t are called e and i.



(A) SECTION OF TRANSMISSION LINE



(B) TRAVELING WAVES

FIG. 3. Voltage and current waves of arbitrary shape on a lossless transmission line.

Let O' be a moving origin that travels with the same velocity as the wave, and let z be a distance along the line, measured to the left from the moving origin. The shapes of the traveling waves of voltage and current may then be specified by writing e and i as functions of z, thus:

$$e = e(z) \qquad \text{and} \qquad i = i(z)$$

If in Fig. 3 the moving origin O' coincides at time $t = 0$ with O, the fixed origin of s, then the distance OO' is vt and z is equal to $vt - s$. Thus the expressions for voltage and current become

$$e = e(vt - s) \tag{3}$$
$$i = i(vt - s) \tag{4}$$

Equations 3 and 4 may be thought of as a mathematical way of stating that waves of e and i travel at velocity v without change of shape or amplitude. Equation 3, for example, indicates that, if the voltage distribution on the line is examined at time t_1 and again at time $t_1 + \Delta t$, the two distributions are identical except for a shift in position of the pattern by the amount $v\Delta t$. This statement is true because, according to Eq. 3, changing t by the amount Δt and simultaneously changing s by the amount $v\Delta t$ does not alter e, because the change leaves $vt - s$ unchanged.

Let Δs in Fig. 3 be an increment of line length so short that the slope of the curve of e as a function of s is essentially constant throughout the increment. The change of e over the length of the increment is therefore approximately

$$\Delta e = \frac{\partial e}{\partial s} \Delta s$$

where the partial derivative $\partial e/\partial s$ is used to indicate the rate of change of e with s for a constant value of t. The voltage difference Δe is the result of a changing current in the series inductance of the incremental length of line. If l is the inductance per unit length of the line, $l\Delta s$ is the inductance of the incremental length Δs, and

$$\Delta e = -l\Delta s \frac{\partial i}{\partial t}$$

The minus sign is introduced because, for the reference directions of e and i indicated in Fig. 3, Δe is negative when i is increasing. Combining the two expressions for Δe yields

$$\frac{\partial e}{\partial s} = -l \frac{\partial i}{\partial t} \tag{5}$$

In a similar way the current change in the distance Δs can be expressed as $(\partial i/\partial s)\Delta s$ or as $-(c\Delta s)\partial e/\partial t$, where c is the capacitance per unit length of line, and thus

$$\frac{\partial i}{\partial s} = -c \frac{\partial e}{\partial t} \tag{6}$$

Equations 5 and 6 are the partial differential equations for waves on a lossless transmission line. From them, the velocity of the waves and the ratio of e to i can be determined simply if the partial derivatives with respect to s and t are first replaced by total derivatives with respect to z. From Fig. 3, it is evident that

$$\frac{\partial e}{\partial s} = -\frac{de}{dz} \qquad \text{and} \qquad \frac{\partial i}{\partial s} = -\frac{di}{dz}$$

Also, because the waveforms of e and i are moving past point P with velocity v and without change of shape, the time rates at which e and i change are equal to the slopes of the $e(z)$ and $i(z)$ curves multiplied by v; that is,

$$\frac{\partial e}{\partial t} = v\frac{de}{dz} \qquad \text{and} \qquad \frac{\partial i}{\partial t} = v\frac{di}{dz}$$

By means of these relations, Eqs. 5 and 6 are changed to

$$\frac{de}{dz} = lv\frac{di}{dz} \tag{7}$$

$$\frac{di}{dz} = cv\frac{de}{dz} \tag{8}$$

From the product of Eqs. 7 and 8, the value $1/\sqrt{lc}$ for v is obtained, and from the quotient of the equations the relation $de = \sqrt{l/c}\, di$ is obtained. Integration yields

$$e = \sqrt{\frac{l}{c}}\, i + \text{const}$$

and thus $R_c = \sqrt{l/c}$, the constant of integration being zero if the line is initially uncharged.

Because sine waves are of special interest in high-frequency systems, the general function $e(vt - s)$ in Eq. 3 may be replaced by the sine function

$$e = E_m \sin \beta(vt - s) \tag{9}$$

where E_m is the peak value of the sine wave of voltage and β is a scale factor for the axis of z or $vt - s$ and determines the extent to which the sine wave is spread out along the line. Equation 9 may be written as

$$e = E_m \sin (\beta vt - \beta s)$$

In this form βv is seen to specify the *time rate of change of phase* of the sine wave for fixed values of s, and thus βv equals the angular frequency ω of the sine wave; that is,

$$\beta v = \omega = 2\pi f \tag{10}$$

Similarly, β specifies the *rate of change of phase with distance* for fixed values of t and is called the *phase constant* of the wave. The *period* of the wave T is the *time* required for the phase to change through 1 cycle or 2π radians, and thus $\omega T = 2\pi$. Similarly, the *wavelength* λ is the *distance* in which the phase changes by 2π radians, or $\beta\lambda = 2\pi$, and

$$\lambda = \frac{2\pi}{\beta} \tag{11}$$

Combining Eqs. 10 and 11 yields the relation given in Chap. I (Eq. 8, p. 32),

$$v = f\lambda \tag{12}$$

Rewriting Eq. 9 in terms of ω and β and writing also the equation for the corresponding current wave yield

$$e = E_m \sin (\omega t - \beta s) \tag{13}$$

$$i = \frac{E_m}{R_c} \sin (\omega t - \beta s) \tag{14}$$

This pair of equations will be used as the basic expressions for sinusoidal traveling waves of voltage and current on a lossless transmission line.

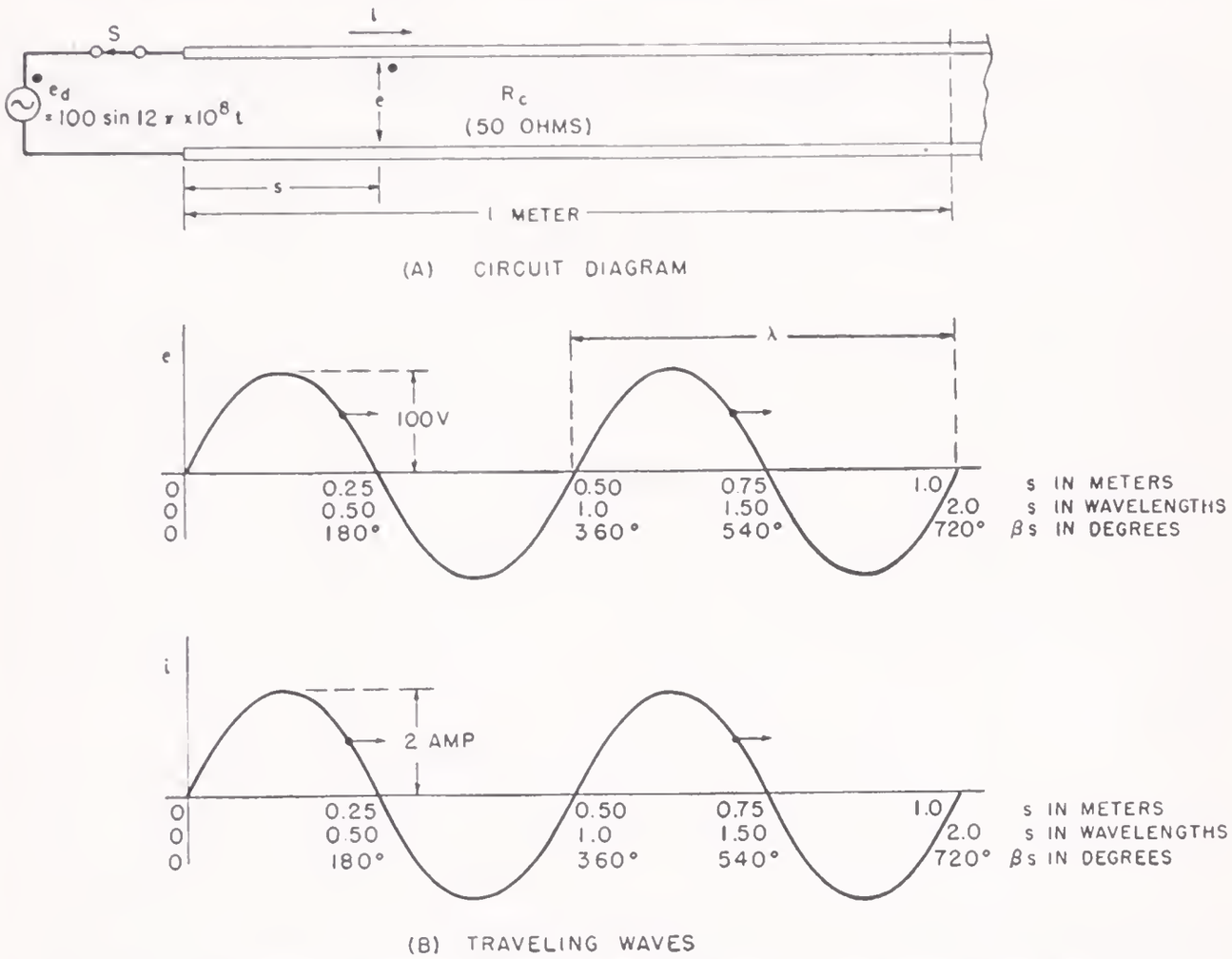


FIG. 4. Traveling sine waves of voltage and current on a lossless transmission line.

As an example of sinusoidal voltage and current relations on a lossless line, consider in Fig. 4A that a voltage of the form $e_d = 100 \sin 12\pi \times 10^8 t$ (peak value of 100 volts and frequency of 600 Mcps) is applied through switch S to an infinitely long lossless line, and assume that the characteristic resistance of the line is 50 ohms. At the instant the switch is closed, a sinusoidal voltage wave having a peak value of 100 volts travels away from the input terminals. In phase with this wave is a current wave of peak value

$$\frac{E_m}{R_c} = \frac{100}{50} = 2 \text{ amp}$$

If the waves are assumed to be traveling at the speed of light (3×10^8 m per sec), the phase constant is

$$\beta = \frac{\omega}{v} = \frac{12\pi \times 10^8}{3 \times 10^8} = 4\pi \text{ radians per m}$$

and the wavelength λ is $2\pi/\beta = 0.5$ m. Thus 2 cycles of the waves occur in the 1-m section as shown in Fig. 4*B*. The relations for voltage and current at any distance s from the input terminals may be written from Eqs. 13 and 14 as

$$e = 100 \sin (12\pi \times 10^8 t - 4\pi s)$$
$$i = 2 \sin (12\pi \times 10^8 t - 4\pi s)$$

Actual lines, of course, are not infinite in length, as assumed in the example of Fig. 4, but rather extend a definite distance from the generator to a load. If the impedance of the load is the characteristic resistance of the line, the load has the same effect upon the line as would an additional line section of infinite length. Thus voltages and currents in the actual line are the same as those in the initial portion of an infinite line. If the load has a different impedance, reflections occur at the load terminals, and the effects obtained depend upon load impedance in relation to the characteristic resistance and upon the line length in relation to a wavelength. The effects of different types of loads are discussed in Arts. 3 through 6. The length of a line in relation to the wavelength is called the *electrical length* of the line and depends upon the physical length and the frequency of operation. In the example of Fig. 4, the electrical length of the 1-m line section at a frequency of 600 Mcps is two wavelengths, 720 deg, or 4π radians.

In the study of transmission lines, complex numbers are frequently used to represent rms values and phase angles of sinusoidal functions of time in the same way as in the analysis of other electric circuits (see Arts. 5 and 6, Chap. VI). For example, in Fig. 5 the generator voltage e_d of peak value 100 volts may be taken as the reference wave and represented by the complex number

$$E_d = \frac{100}{\sqrt{2}} \epsilon^{j0^\circ} = 70.7\epsilon^{j0^\circ}$$

At a point 90 electrical degrees from the input terminals, the voltage wave has the same amplitude as E_d but is delayed 90 deg in phase and is represented by the complex number $70.7\epsilon^{-j90^\circ}$. Complex numbers standing for voltage and current at quarter-wave intervals along the line are indicated graphically in Fig. 5*B*.

In more general terms, the complex-number relation between the voltage of a traveling wave at any point of a lossless transmission line and the

voltage at the input terminals is

$$E = E_d \epsilon^{-j\beta s} \tag{15}$$

and the current relation is

$$I = \frac{E_d}{R_c} \epsilon^{-j\beta s} \tag{16}$$

where E and I are complex numbers representing the voltage and current at a distance s from the input terminals of the line and E_d and I_d are the values of E and I at the input terminals (where $s = 0$). Equations 15 and 16 are the complex-number expressions for a traveling wave of voltage

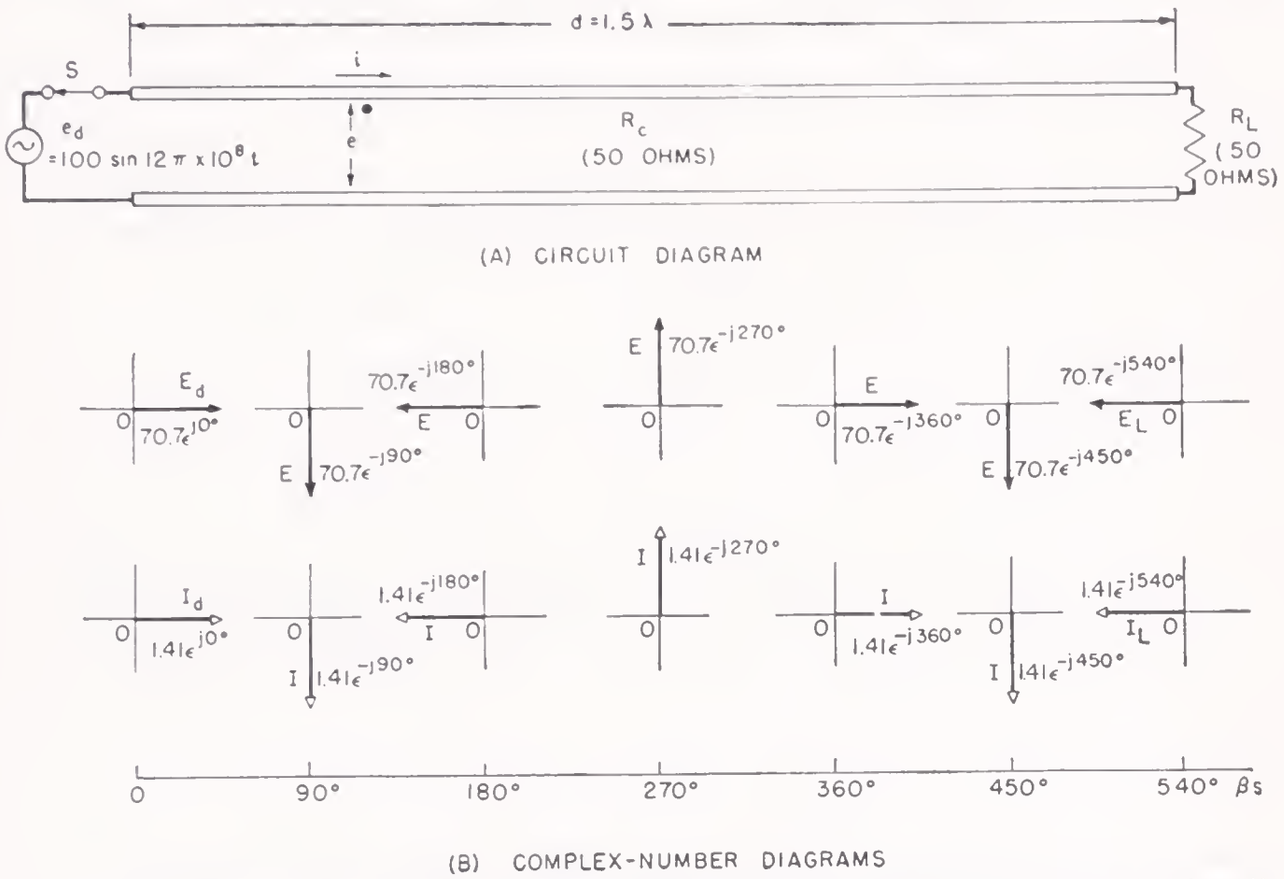
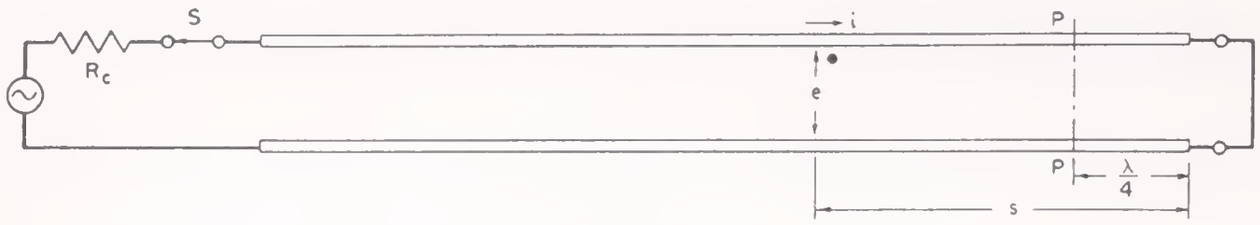


FIG. 5. Complex-number diagrams representing voltage and current along a lossless line terminated in its characteristic resistance.

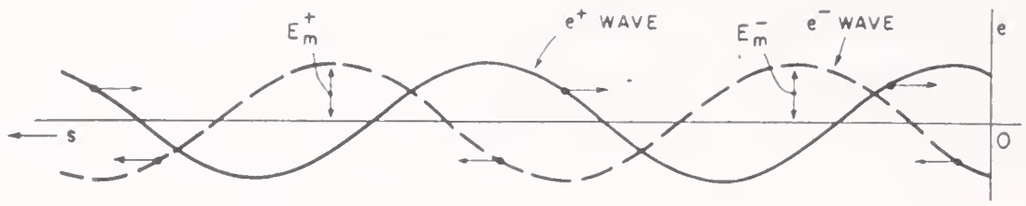
and current that contain the same information about the phase and amplitude of the waves as do the time-function expressions of Eqs. 13 and 14.

3. Lines Terminated in a Short Circuit. When a line is terminated in an impedance different from R_c, the voltage and current on the line are no longer the result of a single wave traveling from generator to load. Instead, the total voltage and current are the sum of two waves traveling in opposite directions.

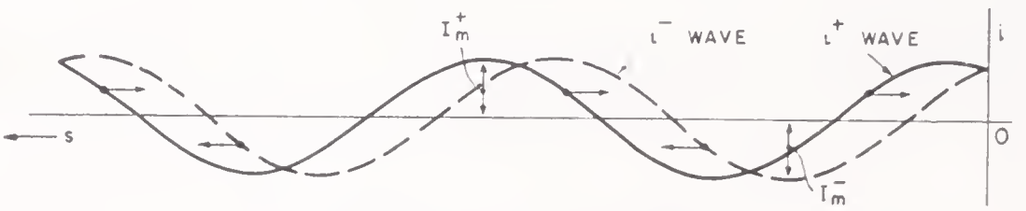
As an example, consider the diagram of Fig. 6*A* in which a sine-wave generator with an internal resistance R_c is connected through switch S to a line terminated in a short circuit. The behavior of the line after the switch is closed is similar to that of a shorted-end line with a direct voltage



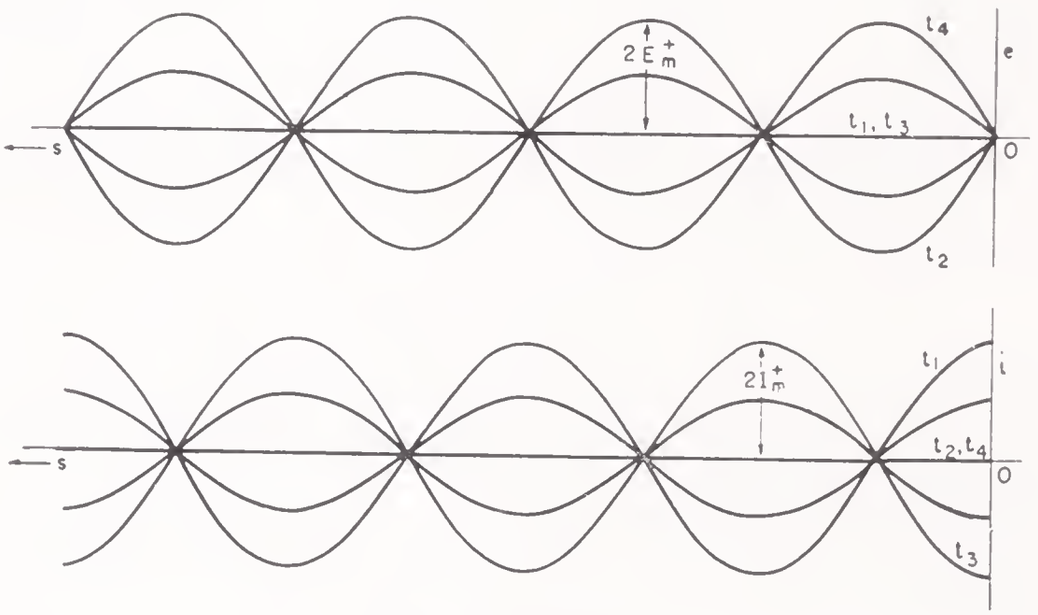
(A) CIRCUIT DIAGRAM



(B) TRAVELING WAVES OF VOLTAGE



(C) TRAVELING WAVES OF CURRENT



(D) STANDING WAVES

FIG. 6. Voltage and current waves on a shorted-end line.

applied (see Art. 24, Chap. II), except, of course, that sinusoidal waves of voltage and current instead of step waves travel along the line. Closing the switch causes in-phase waves of voltage and current to travel to the short circuit. In order that the voltage at the short circuit may be zero at all times, the voltage reflected at the short circuit must be equal in magnitude to the incident voltage, but reversed in polarity. As the reflected voltage wave moves toward the input terminals, it is accompanied by a reflected current wave, the ratio of magnitudes being R_c.

Because the generator is assumed to have an internal resistance R_c, the line behaves with respect to the reflected waves exactly like a line terminated in its characteristic resistance. Steady-state conditions are therefore reached upon arrival of the reflected waves at the input terminals. The incident and reflected waves of voltage are drawn for a particular instant in Fig. 6*B*, and the corresponding waves of current are illustrated in Fig. 6*C*.

In Figs. 6*B* and *C*, plus-sign superscripts are used to denote quantities associated with waves traveling toward the load, and minus-sign superscripts refer to quantities associated with waves traveling away from the load. Thus E_m^+ and E_m^- in Fig. 6*B* are, respectively, the peak values of the incident and reflected voltage waves, the values of E_m^+ and E_m^- being equal. Similarly in Fig. 6*C*, the peak values of the current waves are $I_m^+ = E_m^+/R_c$ and $I_m^- = E_m^-/R_c$. The current of the reflected wave would be in phase with the voltage of the reflected wave if the reference direction for the reflected-wave current in the upper conductor were taken toward the generator. However, the reference direction indicated in Fig. 6*A* is used for all current components; therefore the polarity of the reflected current wave is reversed with respect to the reflected voltage wave, and the two current components are in phase at the short circuit.

The total voltage at any point of the line, which is the sum of the two voltage waves in Fig. 6*B*, varies as the component waves move in opposite directions along the line. At time t_1, an instant slightly later than that for which Fig. 6*B* is drawn, the peak of the incident voltage wave is at the short circuit, and the component voltages have equal magnitudes and opposite polarities at each point of the line. Thus the total voltage is everywhere zero, as indicated by the line marked t_1 in Fig. 6*D*. At other instants cancellation of the component waves does not occur. For example, at time t_2, a quarter period later than time t_1, a zero of the component waves in Fig. 6*B* is at the short circuit, and the total voltage (marked t_2 in Fig. 6*D*) is a sine wave of twice the amplitude of either component wave. At time t_3, a quarter period after t_2, cancellation is again obtained; and another quarter period later, at time t_4, a double-amplitude voltage wave of polarity opposite to that for time t_2 is obtained.

The total-voltage pattern of Fig. 6*D* is called a *standing wave*. By performing additions at times intermediate to the quarter-period instants, it may be shown that the total voltage has a sine-wave distribution along the line, with zeros at the short circuit and at half-wave intervals from the short circuit. The amplitude and polarity of this sine wave varies as the voltage at each point changes sinusoidally with time. The points of zero voltage are called voltage *nodes* (or *nodal points*), and the points of maximum voltage midway between the nodes are *antinodes*.

Point-by-point additions of the component current waves in Fig. 6*C*

may be made in similar fashion. The total-current pattern obtained is the standing wave indicated in Fig. 6*D*. The current standing wave differs from the voltage standing wave only in position along the line—the nodes of the current wave occur at odd quarter-wave distances from the short circuit.

The addition of the incident and reflected waves to form a standing wave may also be performed with the aid of complex numbers. For the shorted-end line of Fig. 7*A*, let E_L^+ and E_L^- be complex numbers representing the voltage of the incident and reflected waves at the short circuit. Because the total voltage at the short circuit must be zero, E_L^- is equal to $-E_L^+$. Let E^+ and E^- represent the incident-wave and reflected-wave voltages at point P located a distance s from the short circuit. In traveling from point P to the short circuit, the incident wave suffers a phase delay βs; in returning to point P, the reflected wave is delayed an equal amount. Thus $E^+ = E_L^+ \epsilon^{j\beta s}$ and $E^- = E_L^- \epsilon^{-j\beta s}$. The total voltage at point P is therefore

$$E = E^+ + E^- = E_L^+ \epsilon^{j\beta s} + E_L^- \epsilon^{-j\beta s} = E_L^+ (\epsilon^{j\beta s} - \epsilon^{-j\beta s})$$

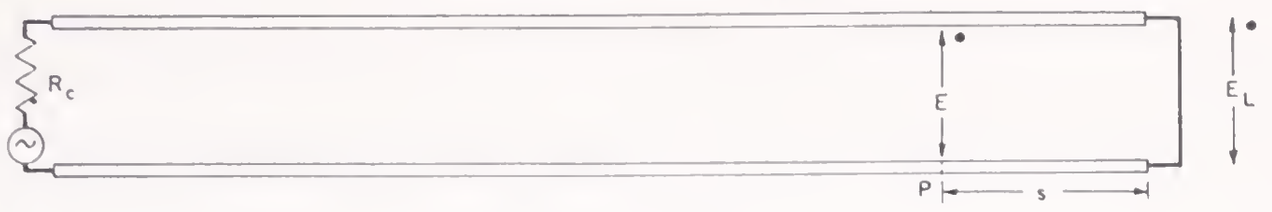
or, because $\sin \beta s = (\epsilon^{j\beta s} - \epsilon^{-j\beta s})/2j$,

$$E = j2E_L^+ \sin \beta s \tag{17}$$

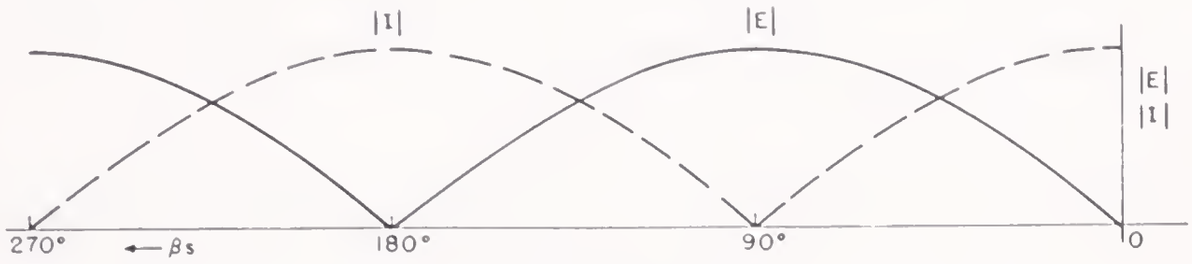
Equation 17 indicates that $|E|$, the rms value of the line voltage, is proportional to the magnitude of $\sin \beta s$. Thus the plot of $|E|$ in Fig. 7*B* takes the form of a series of half cycles of sine waves, the zeros of the pattern occurring at the nodes of the standing wave of voltage. The phase of the line voltage E relative to the phase of E_L^+ is also indicated by Eq. 17. In the half-wave interval adjacent to the short circuit, E leads E_L^+ by 90 deg because of the operator j. The phase of E changes abruptly by 180 deg at each of the nodal points because of the change of sign of $\sin \beta s$ in Eq. 17. Thus the phase variations of E are as indicated by the solid line in Fig. 7*C*.

The addition made in obtaining Eq. 17 is presented graphically by the complex-number diagrams comprising the upper half of Fig. 7*D*. Voltage E_L^+ is used as a reference voltage in these diagrams. Observe that in each diagram the horizontal components of E^+ and E^- cancel, so that E always lags or leads E_L^+ by 90 deg, and $|E|$ changes from zero to $2|E_L^+|$ as E^+ and E^- change from horizontal to vertical positions in the diagrams.

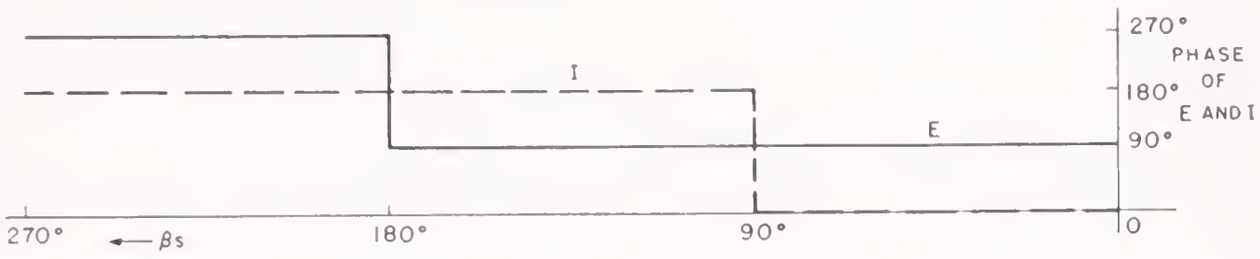
In Figs. 7*B* and *C*, magnitude and phase curves for the standing wave of current are shown by dashed lines, and complex-number diagrams from which these curves may be derived are included in Fig. 7*D*. Alternatively, the curves may be plotted from an equation analogous to Eq. 17. If I^+ and I^- are the complex numbers representing current components at any point of the line and I_L^+ and I_L^- the values of I^+ and I^- at the



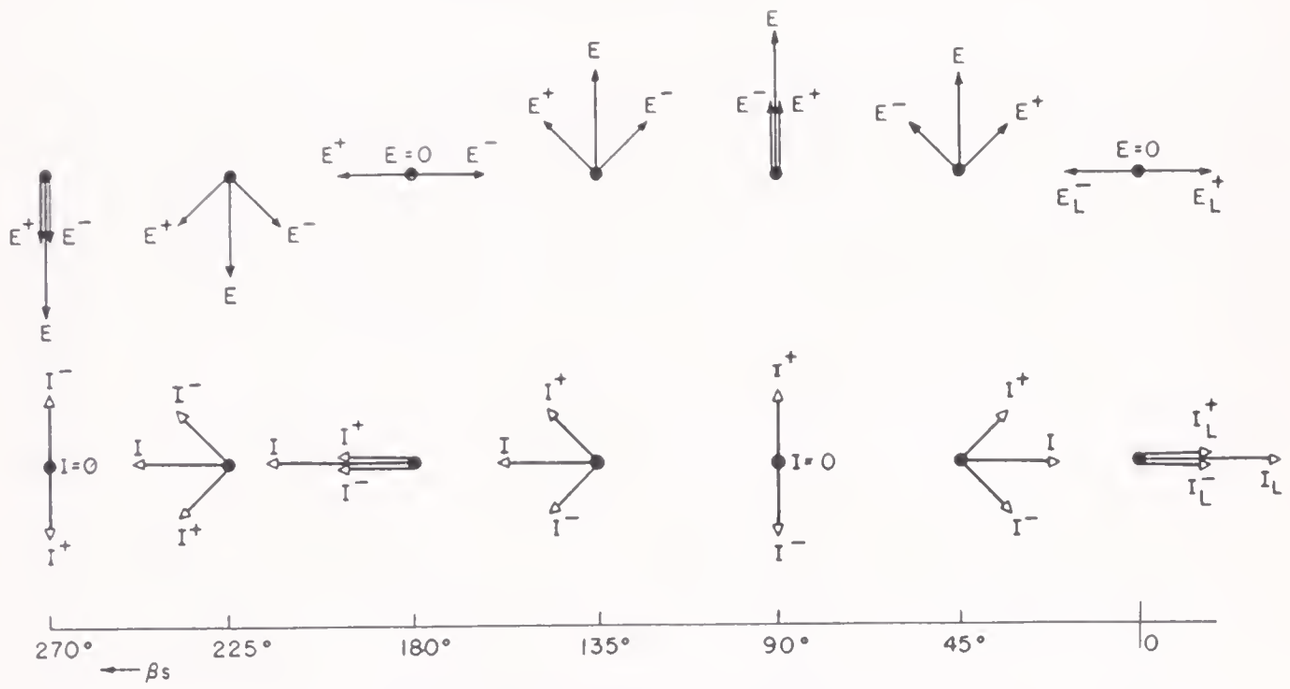
(A) CIRCUIT DIAGRAM



(B) MAGNITUDE OF E AND I



(C) PHASE OF E AND I



(D) COMPLEX-NUMBER DIAGRAMS

FIG. 7. Variations in the magnitude and phase of voltage and current on a shorted-end line.

short circuit, then $I^+ = I_L{}^+\epsilon^{j\beta s}$, $I^- = I_L{}^-\epsilon^{-j\beta s}$, and $I_L{}^+ = I_L{}^-$. Thus the total current I is

$$I = I^+ + I^- = I_L{}^+(\epsilon^{j\beta s} + \epsilon^{-j\beta s}) = 2I_L{}^+ \cos \beta s$$

or

$$I = 2\frac{E_L{}^+}{R_c} \cos \beta s \tag{18}$$

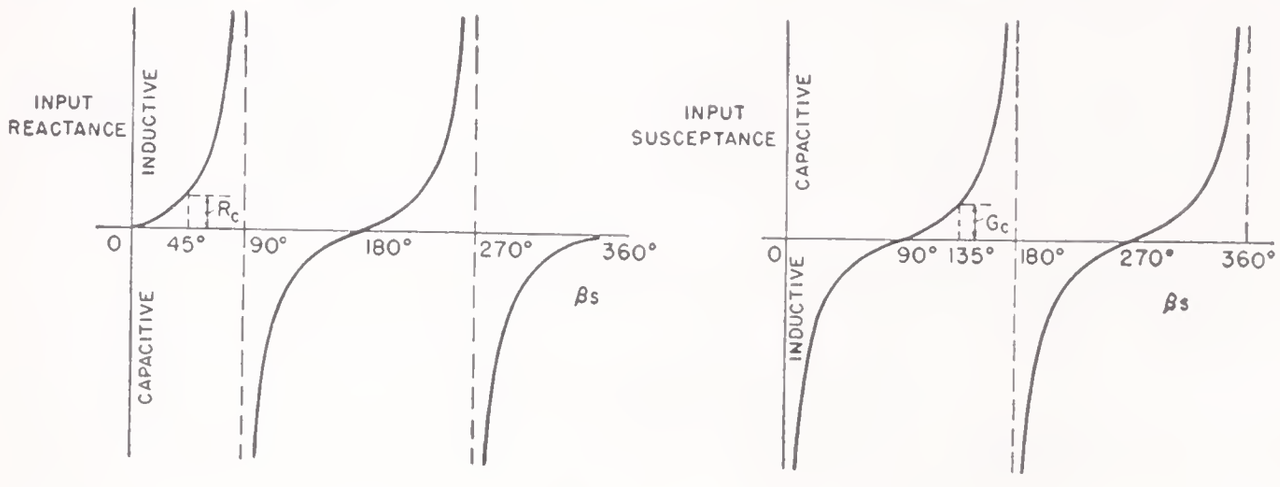
Very important in the use of transmission lines is the input impedance of a section of line. Input impedance is defined as the ratio of the complex numbers representing total voltage and total current at the input terminals of the section. If the line is lossless, the impedance magnitude may have any value between zero and infinity (depending upon the nature of the load impedance and the length of the line), and the impedance angle may have any value between +90 and −90 deg. Thus the input impedance of a lossless section of line may be a pure resistance, a pure reactance, or a combination of both.

The general nature of the input impedance of a section of shorted-end line may be determined from Fig. 7*C*. Because of the 90-deg phase relation between voltage and current, the input impedance is purely reactive. For line lengths less than a quarter wavelength, I lags E, and the reactance is inductive. For line lengths between a quarter and a half wavelength, I leads E and the reactance is capacitive. The cycle of values is repeated in each succeeding half-wave interval.

A relation for the input impedance Z of a shorted-end line as a function of line length may be obtained by taking the ratio of E to I in Eqs. 17 and 18. Thus

$$Z = \frac{E}{I} = jR_c \tan \beta s \tag{19}$$

The purely reactive nature of the input impedance is indicated by the presence of only an imaginary component in the complex input-impedance expression of Eq. 19. The magnitude of the reactance depends upon the characteristic resistance of the line and varies with line length in the manner of a tangent function. Reactance magnitudes of various line



(A) REACTANCE CURVE (B) SUSCEPTANCE CURVE

FIG. 8. Input reactance and susceptance curves for a shorted-end line.

lengths are shown graphically in Fig. 8*A*. Observe that because Z is a periodic function of βs, the period being 180 deg, a change in line length of any integral number of half wavelengths has no effect upon the input

impedance. It is also important to note that, because the phase constant β is proportional to frequency, the input-impedance variations of Fig. 8*A* may be obtained either by varying the frequency of the voltage applied to a fixed length of line or by varying the physical length of the line at a fixed frequency.

The input admittance of a section of line is the reciprocal of the input impedance and for a shorted-end line is given by the expression

$$Y = \frac{1}{Z} = \frac{1}{jR_c \tan \beta s} = -jG_c \cot \beta s \tag{20}$$

where $G_c = 1/R_c$ is the characteristic conductance of the line. The input admittance is therefore a susceptance which varies with line length in the manner of a cotangent function. Variations in input susceptance with line length are represented graphically in Fig. 8*B*.

4. Lines Terminated in an Open Circuit and in a Reactance. The voltage and current distribution along an open-end line may be deduced in much the same manner as for the shorted-end line, the only difference being that, at the open end, the phase relations of the incident and reflected waves are such as to make the current rather than the voltage zero. A quicker method is to consider that the line of Fig. 6*A* is composed of two sections joined at point P. If the section to the right of P is a quarter wave in length, its input impedance is infinite and no current flows at the junction of the two sections. Hence, cutting the conductors at P has no effect on the voltage and current along the line to the left of P, and the effect of an open-circuit termination may be obtained by removing the quarter-wave section. The standing-wave patterns of voltage and current for the open-end section may therefore be derived directly from those of Fig. 6*D* by moving the origin a quarter wavelength to the left. The resultant voltage and current patterns for the open-end line of Fig. 9*A* are shown in Fig. 9*B*.

The complex-number diagrams at the quarter-wave point of the shorted-end line in Fig. 7 also represent voltage and current relations at the end of the open line. If E_L^+ and E_L^- represent the incident and reflected voltage waves at the open circuit, the total voltage E at any point of the line is

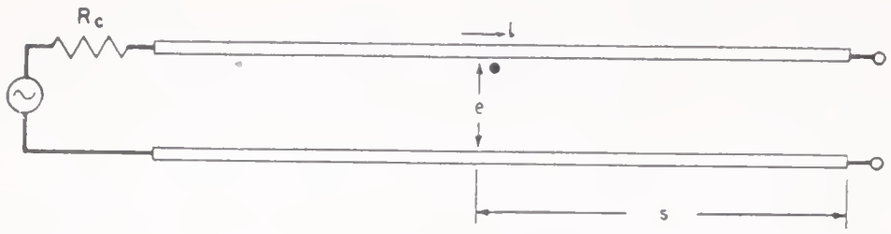
$$E = E_L^+ \epsilon^{j\beta s} + E_L^- \epsilon^{-j\beta s}$$

where s is the distance of the point from the open end, or because $E_L^+ = E_L^-$ at the open circuit,

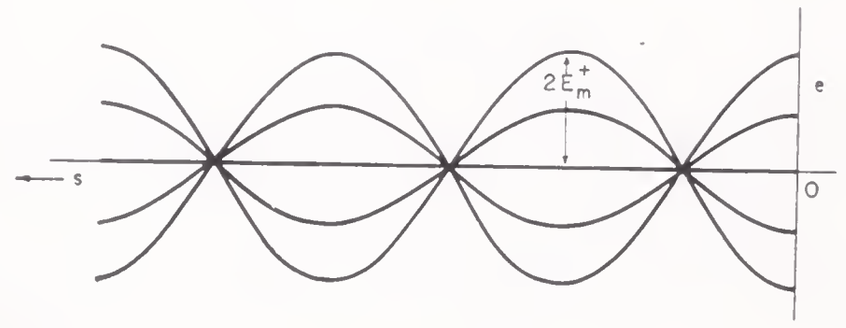
$$E = 2E_L^+ \cos \beta s \tag{21}$$

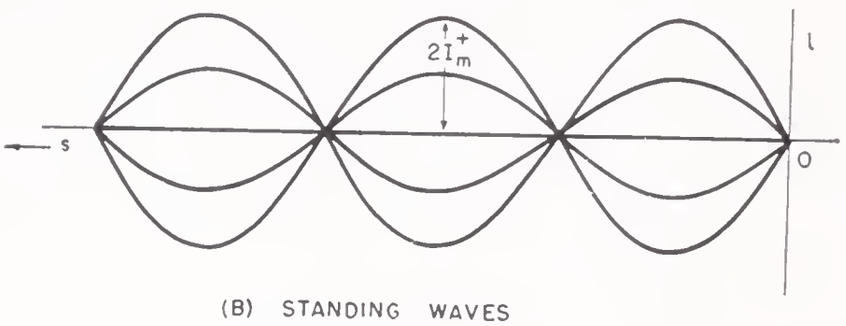
Similarly, the current at any distance from the open circuit is

$$I = I_L^+ \epsilon^{j\beta s} + I_L^- \epsilon^{-j\beta s}$$

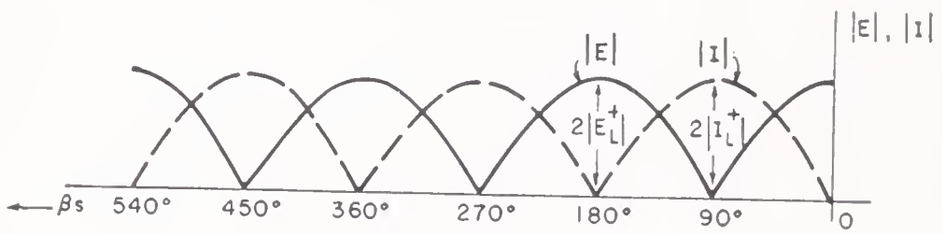


(A) CIRCUIT DIAGRAM

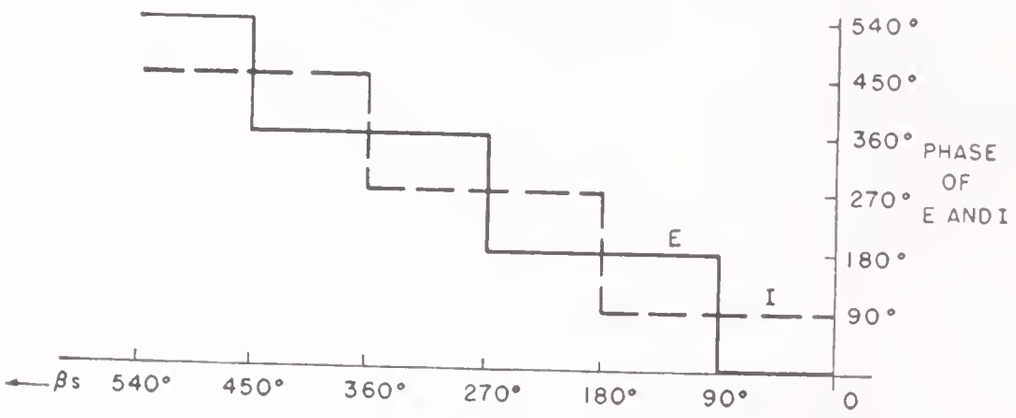




(B) STANDING WAVES



(C) MAGNITUDE OF E AND I



(D) PHASE OF E AND I

FIG. 9. Voltage and current relations for an open-end line.

At the open circuit

$$I_L^- = -I_L^+ = \frac{-E_L^+}{R_c}$$

and therefore

$$I = j2\frac{E_L^+}{R_c}\sin \beta s \tag{22}$$

Variations in the magnitude and phase of E and I along the line are plotted in Figs. 9C and D.

The manner in which the input impedance of an open-end section of line varies with the length of the section may be obtained from the impedance curve of the shorted-end line (Fig. 8A) by transferring the origin to the point $\beta s = 90°$. Thus for the open line the input reactance is capacitive for lengths less than a quarter wave and inductive for lengths between a quarter wave and a half wave. Dividing Eq. 21 by Eq. 22 gives as the input-impedance relation for the open-end line

$$Z = -jR_c \cot \beta s \tag{23}$$

The corresponding input admittance relation is

$$Y = \frac{1}{Z} = jG_c \tan \beta s \tag{24}$$

The effect on the standing-wave pattern of terminating a line in a pure reactance may also be determined from the shorted-end line. For

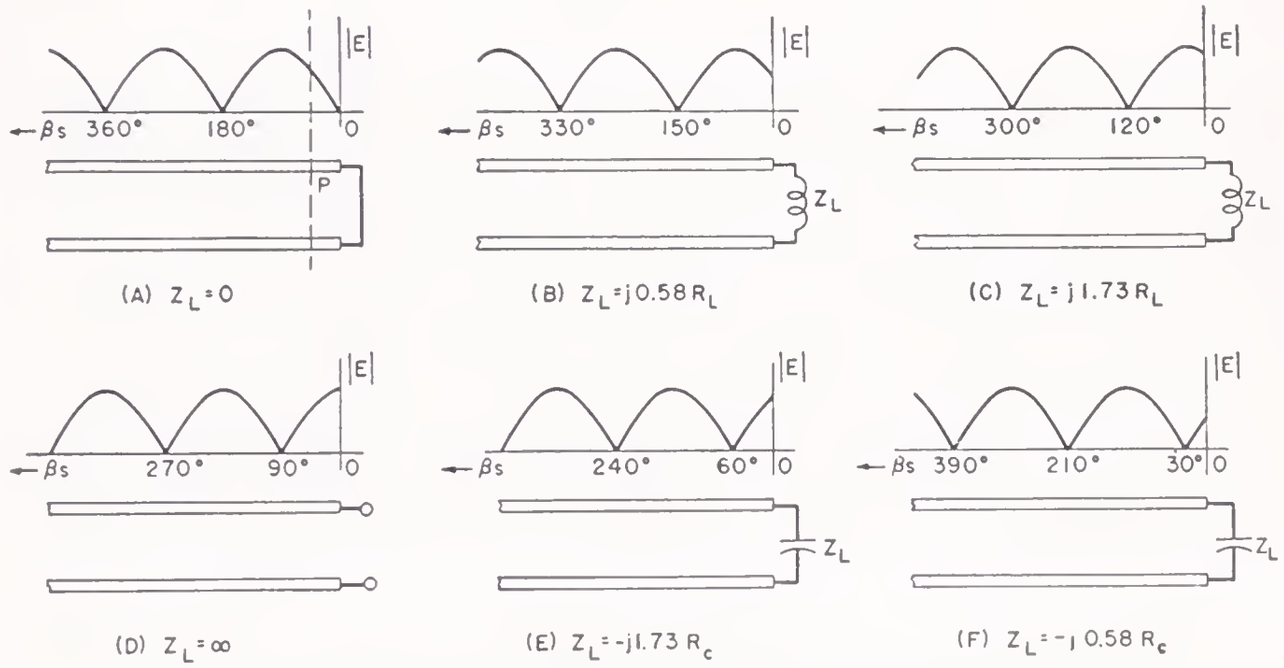


FIG. 10. Standing-wave patterns for lines terminated in a reactance.

example, if point P in Fig. 10A is less than a quarter wave from the short circuit, the input impedance of the line section to the right of P is an inductive reactance. Therefore, an inductor of equal reactance may be

substituted for the shorted-end section without changing the voltage and current values on the rest of the line. The standing-wave patterns for the lines of Figs. 10*B* and *C* are obtained in this fashion, and the remaining parts of Fig. 10 are obtained in a similar way for positions of point *P* in Fig. 10*A* farther from the short circuit. The open-circuit termination of Fig. 10*D* applies for *P* a quarter wave from the short circuit, and the capacitive terminations of Figs. 10*E* and *F* correspond to a position of *P* more than a quarter wave but less than a half wave from the short circuit.

Observe that the inductive or capacitive terminating reactances of Fig. 10 can be considered equivalent to an additional line section of fixed length at only a single frequency. The variation of reactance with frequency for a transmission-line section is different from the reactance-frequency curve of a lumped-parameter element. Although the standing-wave patterns on the line of Fig. 10*B* and the line section to the left of point *P* in Fig. 10*A* may be alike at one frequency, the patterns change in a different manner when the frequency is changed.

5. Lines Terminated in a Resistance. The behavior of traveling step waves of voltage on a lossless line terminated in a resistance is discussed in Art. 24, Chap. II. The incident wave is reflected at the load resistance with an amplitude and polarity determined by the value of the load resistance R_L relative to R_c, the characteristic resistance of the line. The ratio of the reflected voltage wave to the incident voltage at the load is given by the load reflection coefficient

$$\Gamma_L = \frac{R_L - R_c}{R_L + R_c} \tag{25}$$

Because a wave of any shape can be approximated by the sum of many small step waves, as in Art. 2 (see Fig. 2, p. 472), the reflection coefficient can be applied to determine instantaneous values of waves reflected from a resistance no matter what the shape of the incident wave. In particular, for an incident sine wave, calculation of instantaneous values shows that the reflected wave is also sinusoidal and of an amplitude related to the amplitude of the incident wave by the magnitude of Γ_L. At the load, the reflected wave is in phase with the incident wave if Γ_L is positive (R_L greater than R_c) or 180 deg out of phase if Γ_L is negative (R_L less than R_c). In terms of complex numbers, the relation is

$$\frac{E_L^-}{E_L^+} = \Gamma_L \tag{26}$$

where E_L^- and E_L^+ are complex numbers representing, respectively, the incident and reflected waves of voltage at the load and Γ_L may be considered to be a complex number having an amplitude between zero and unity and a phase angle of zero or 180 deg.

As an example of the way in which voltage and current relations on a line depend on Γ_L, consider the circuit of Fig. 11*A*. A generator of internal resistance R_c applies a sine wave of voltage to the line, which is terminated by a resistance $R_L = 3R_c$. Because the line faces its characteristic resistance at the generator, steady-state conditions are reached after one round trip of the initial traveling waves. Voltage relations at the load in the steady state are illustrated by the upper-right complex-number diagram of Fig. 11*B*. The incident voltage wave at the load (represented by E_L^+) is taken as the reference wave. The line marked E_L^- is drawn in phase with E_L^+ and with half the amplitude of E_L^+, because for the value of R_L assumed, $\Gamma_L = 0.5$. The total load voltage is the sum of the incident and reflected voltages and is therefore $1.5E_L^+$.

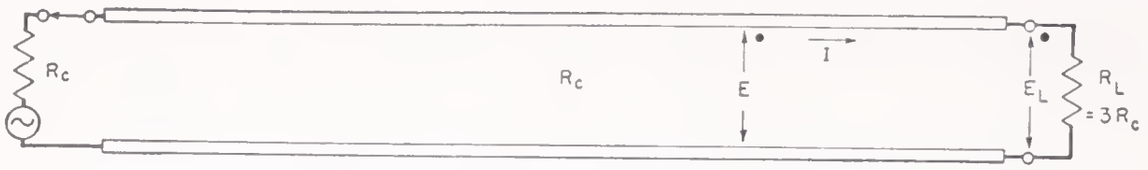
The corresponding diagram for load current is also shown in Fig. 11*B*. The incident- and reflected-current components are obtained from the relations $I_L^+ = E_L^+/R_c$ and

$$I_L^- = -\frac{E_L^-}{R_c} = -\Gamma_L \frac{E_L^+}{R_c}$$

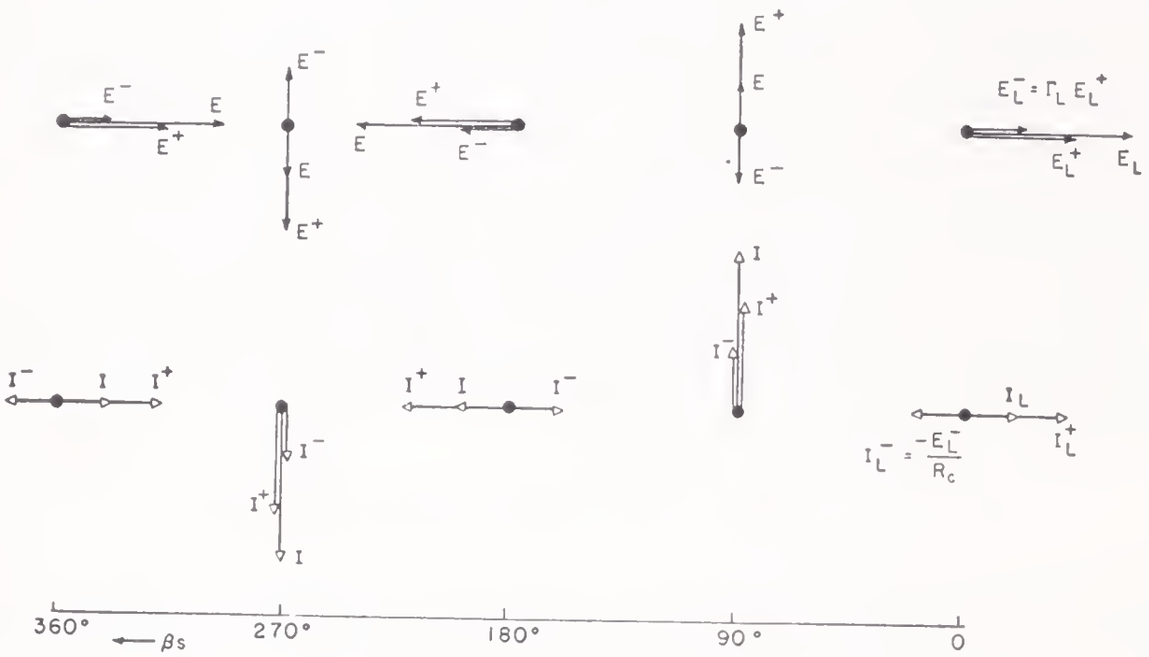
The total current at the load is the sum of I_L^+ and I_L^- and is therefore $0.5E_L^+/R_c$. Observe that the ratio E_L/I_L is $3R_c$ as it must be because of the value assumed for R_L.

At a distance s from the load terminals, the incident voltage and current waves are advanced in phase through an angle βs and the reflected waves are retarded through an equal angle. Incident and reflected waves of voltage and current, as well as the total voltage and current, are represented for several points of the line by the complex-number diagrams of Fig. 11*B*. The magnitude and phase variations of total voltage and current are plotted in Figs. 11*C* and *D*. Values for plotting these curves may be obtained from the complex-number diagrams or, analytically, by adding the complex-number expressions for the incident and reflected voltage at each point of the line. Observe that, although the magnitude of the total voltage varies with position on the line, the variations are not so great as with total reflection (Fig. 9*C*, for example), because the reflected wave is smaller than the incident wave. The phase variation is intermediate to the phase variation indicated in Fig. 9*D* and to the uniform change of phase with distance associated with a single traveling wave.

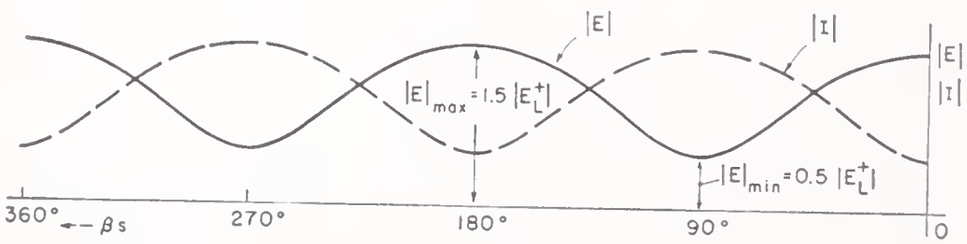
The curves of Fig. 12 are drawn to show the effect on the standing-wave pattern of change in the load resistance. Figures 12*A* and *B* apply for values of R_L greater than R_c and show that, as R_L is made more nearly equal to R_c, the position of the pattern is unchanged but the peaks and troughs become less pronounced; that is, the magnitude of the total voltage at each point of the line becomes more nearly that of the incident-



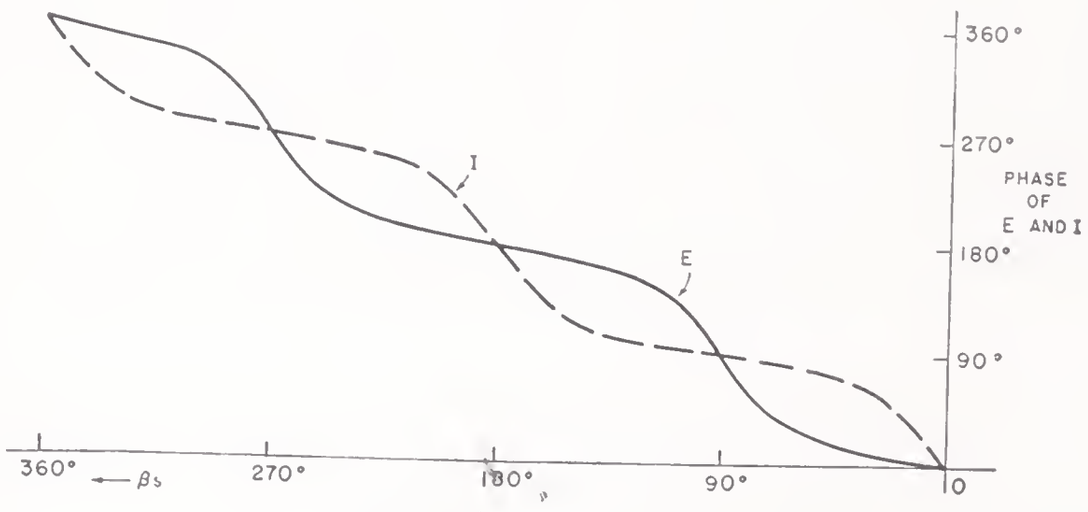
(A) CIRCUIT DIAGRAM



(B) COMPLEX-NUMBER DIAGRAMS



(C) MAGNITUDE OF E AND I



(D) PHASE OF E AND I

FIG. 11. Voltage and current relations for a line terminated in a resistance $R_L = 3R_c$.

voltage wave. Exact equality is reached for a value of R_L equal to R_c (see Fig. 12*C*). Making R_L less than R_c causes an interchange in the positions of peaks and troughs, as indicated in Fig. 12*D*. As R_L decreases, the variations of voltage become greater until for $R_L = 0$ (Fig. 12*E*) the pattern for a shorted-end line described in Art. 3 is obtained.

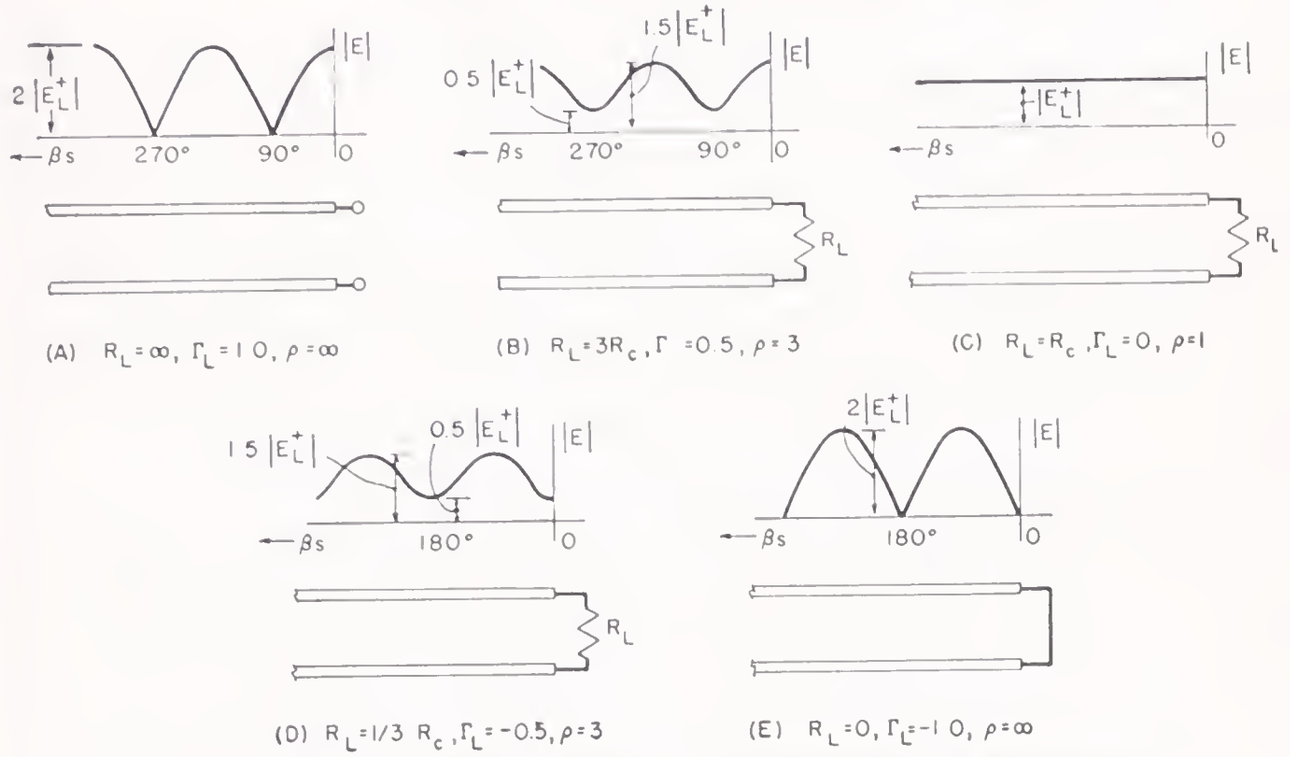


FIG. 12. Standing-wave patterns for lines terminated in a resistance.

A quantity which describes the amount of variation of the rms voltage is the *standing-wave ratio*. It is denoted by the symbol ρ and is defined as the ratio of the maximum rms voltage to the minimum rms voltage on the line. If these voltages are designated as $|E|_{\max}$ and $|E|_{\min}$, respectively, the standing-wave ratio ρ is

$$\rho = \frac{|E|_{\max}}{|E|_{\min}} \tag{27}$$

The standing-wave patterns of Fig. 12 indicate that lines having the same magnitude of load reflection coefficient have equal values of standing-wave ratio. If $|\Gamma_L| = 1$, for example, corresponding to an open- or short-circuit termination, $|E|_{\min}$ is zero and the standing-wave ratio is infinity. If $|\Gamma_L| = 0.5$, corresponding to load resistances of $3R_c$ or $0.33R_c$, the value of ρ is $1.5E_L{}^+/0.5E_L{}^+$, or 3. If $R_L = R_c$, the reflection coefficient is zero and $\rho = 1$, because the voltage is uniform over the line. As $|\Gamma_L|$ is varied from 0 to 1, therefore, ρ varies between 1 and infinity.

The relation between ρ and $|\Gamma_L|$ may be derived from the complex-number diagrams of Fig. 11*B* by noting that maximum voltage occurs at points where the incident and reflected waves are in phase and minimum voltage occurs where the two waves are in phase opposition. At a

maximum-voltage point

$$|E|_{\max} = |E_L^+| + |\Gamma_L|\,|E_L^+| = |E_L^+|(1 + |\Gamma_L|) \tag{28}$$

and at a minimum-voltage point

$$|E|_{\min} = |E_L^+| - |\Gamma_L|\,|E_L^+| = |E_L^+|(1 - |\Gamma_L|) \tag{29}$$

Dividing Eq. 28 by Eq. 29 gives

$$\rho = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|} \tag{30}$$

In the measurement of standing-wave ratio, readings proportional to the square of voltage are sometimes obtained, and the ratio of such readings is called the power standing-wave ratio. In this book, standing-wave ratio always refers to the voltage standing-wave ratio ρ rather than to the power ratio ρ^2.

6. Lines Terminated in a Complex Impedance. Waves of voltage and current that result when sinusoidal traveling waves on a transmission line are reflected from a linear terminating impedance are sinusoidal in form and have a phase and amplitude determined by the incident waves and the impedance causing the reflection. In Art. 4, the reflected wave is shown to have the same amplitude as the incident wave if the impedance is a reactance. In Art. 5, the reflected wave is found to have a smaller amplitude and to have, at the load, a phase angle of zero or 180 deg with respect to the incident wave if the impedance is a resistance. In this article, a complex load impedance is considered, and it is shown that (depending upon the value of the impedance) the reflected wave may have any amplitude not greater than that of the incident wave and any phase relative to the incident wave. The phase and amplitude relations between reflected and incident waves at the load are expressed by the complex-number reflection coefficient $\Gamma_L = E_L^-/E_L^+$, the value of Γ_L being determined by the requirement that the total load voltage be the product of the total load current and the load impedance.

The relation between load impedance and reflection coefficient may first be illustrated by an example in which the value of the reflection coefficient is given and the corresponding load-impedance value is to be calculated. In Fig. 13*A* a line connected to a sine-wave generator of internal impedance R_c is terminated in an impedance Z_L that provides a load reflection coefficient Γ_L equal to $0.5\epsilon^{j60°}$. The upper right-hand diagram of Fig. 13*B* relates the incident and reflected components of voltage at the load during steady-state operation, the reflected voltage $E_L^- = \Gamma_L E_L^+$ being drawn 60 deg ahead of E_L^+ in phase and with half the amplitude of E_L^+. The sum of the complex numbers E_L^+ and E_L^- yields

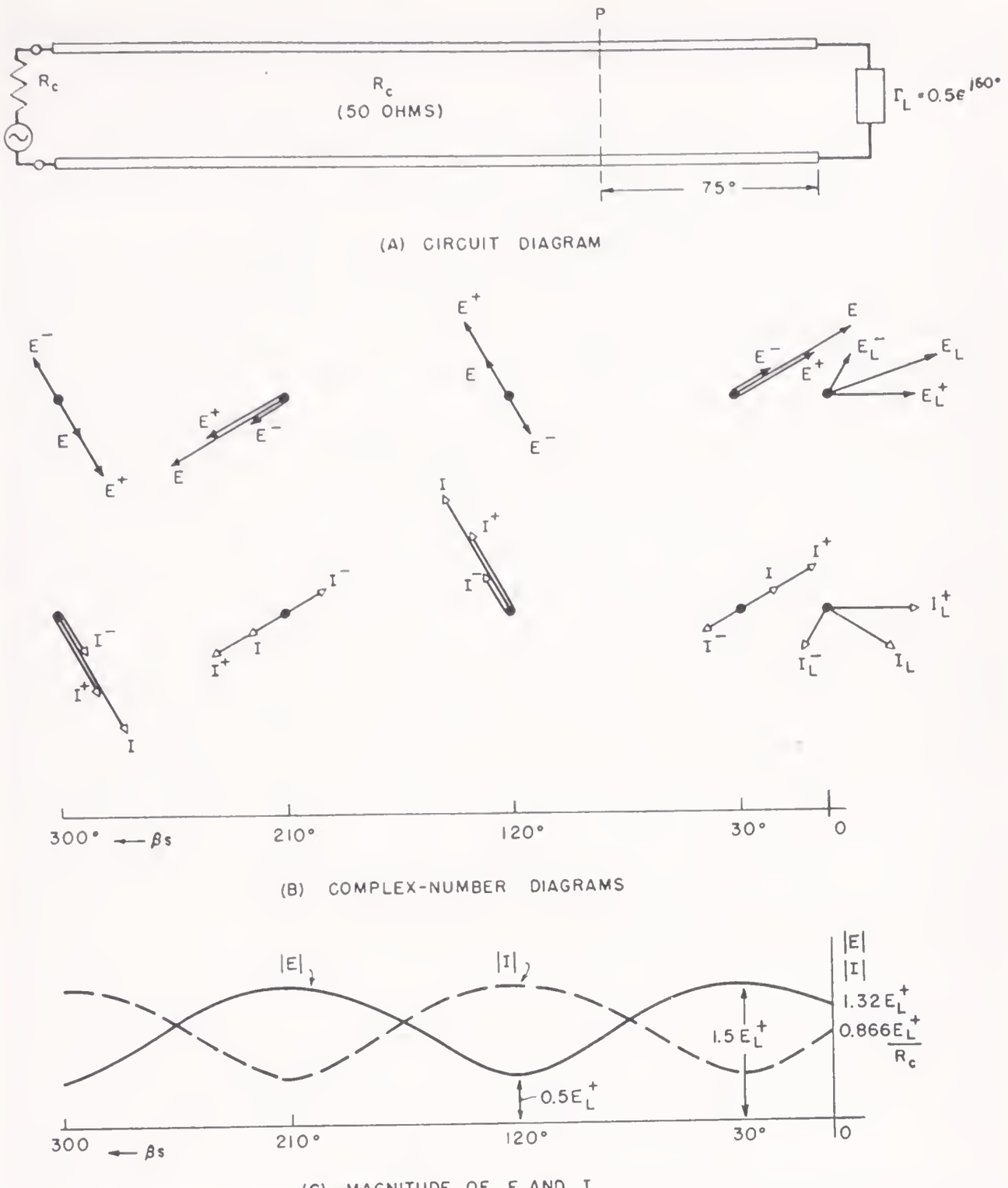


FIG. 13. Voltage and current relations for a line terminated in a complex impedance.

the total load voltage

$$E_L = E_L^+ + 0.5E_L^+\epsilon^{j60°} = E_L^+(1 + 0.5 \cos 60° + j0.5 \sin 60°)$$
$$= 1.32E_L^+\epsilon^{j19.1°}$$

Similarly, the lower right-hand diagram of Fig. 13*B* presents the current relations, and the total load current is

$$I_L = I_L^+ - 0.5I_L^+\epsilon^{j60°} = \frac{E_L^+}{R_c}(1 - 0.5 \cos 60° - j0.5 \sin 60°)$$
$$= 0.866\frac{E_L^+}{R_c}\epsilon^{-j30°}$$

The load impedance that produces the assumed value of reflection coefficient is therefore

$$Z_L = \frac{E_L}{I_L} = \frac{1.32}{0.866} R_c \epsilon^{j(19.1°+30°)} = 1.53 R_c \epsilon^{j49.1°}$$

For a 50-ohm line, Z_L has the value $50 + j57.8$ ohms.

The procedure for determining the remaining diagrams of Fig. 13*B* is the same as that used with the resistance load of Fig. 11*B*. At a distance s from the load, the incident-wave components of voltage and current are advanced in phase by an angle βs, the reflected-wave components are retarded by the same angle, and total voltage and current are determined by adding the incident and reflected components. Data for plotting the curves of voltage and current magnitude in Fig. 13*C* are obtained from these diagrams or from the corresponding complex-number expressions.

Observe that the curves of $|E|$ and $|I|$ in Fig. 13*C* differ from the corresponding curves of Fig. 11*C* only by a shift of position along the axis of βs. The reason for the similar shapes is that $|\Gamma_L|$ is the same for both figures. In general, *the shape of a standing-wave pattern depends upon the magnitude of the load reflection coefficient, and the position of the pattern along the line depends upon the angle of the load reflection coefficient.*

This statement is true because, as may be observed in the construction of the diagrams of Fig. 13*B*, a change in the angle of Γ_L may be compensated by the angle associated with a shift in position along the line. Consider the ratio of the voltage of the reflected wave E^- to the voltage of the incident wave E^+ at any point along the line:

$$\frac{E^-}{E^+} = \frac{E_L{}^- \epsilon^{-j\beta s}}{E_L{}^+ \epsilon^{j\beta s}} = \Gamma_L \epsilon^{-j2\beta s}$$

If Γ_L is written in terms of its magnitude $|\Gamma_L|$ and phase angle ψ,

$$\frac{E^-}{E^+} = |\Gamma_L| \epsilon^{j\psi} \epsilon^{-j2\beta s} = |\Gamma_L| \epsilon^{j(\psi - 2\beta s)} \tag{31}$$

The standing-wave pattern depends upon the relation of E^- to E^+ at each point of the line. Equation 31 shows that a change of ψ is equivalent to changing the value of s associated with each value of the ratio E^-/E^+ and thus to a shift of the pattern along the line. The component voltages E^- and E^+ are in phase when $\psi - 2\beta s = 0$, that is, at an angular distance $\beta s = \psi/2$ from the load. At this point and at half-wave intervals from it, the total voltage of the standing-wave pattern is a maximum.

The magnitude of Γ_L is shown by Eq. 31 to control the shape of the standing-wave pattern. As $|\Gamma_L|$ changes from unity to zero, the pattern changes shape in the manner illustrated by Figs. 12*A*, *B*, and *C*. Because the shape of the pattern is independent of the angle of Γ_L, the relation

between the standing-wave ratio ρ and $|\Gamma_L|$ specified by Eq. 30 is correct for any complex value of terminating impedance (although derived in Art. 5 for a resistance load).

7. Graphical Determination of Reflection Coefficient. Because of the simple relation between reflection coefficient at the load and the standing-wave pattern, a problem of importance in connection with transmission lines is the determination of reflection coefficient when the load impedance is given, and vice versa. A general relation can be obtained by dividing the load voltage (expressed in terms of incident-wave and reflected-wave components) by the load current (also expressed in terms of components). Thus

$$Z_L = \frac{E_L}{I_L} = \frac{E_L^+ + \Gamma_L E_L^+}{I_L^+ - \Gamma_L I_L^-} = \frac{E_L^+}{I_L^+}\frac{1 + \Gamma_L}{1 - \Gamma_L} = R_c \frac{1 + \Gamma_L}{1 - \Gamma_L}$$

or

$$z_L = \frac{1 + \Gamma_L}{1 - \Gamma_L} \tag{32}$$

where z_L is the per-unit load impedance $z_L = Z_L/R_c$. Equation 32 permits a direct calculation of load impedance when the value of reflection coefficient and R_c are known. The inverse calculation is most readily performed with the aid of the following equation, which is obtained by solving Eq. 32 for Γ_L:

$$\Gamma_L = \frac{z_L - 1}{z_L + 1} \tag{33}$$

Although use of Eqs. 32 and 33 is not difficult, the relationship between Γ_L and z_L is of such importance in transmission-line calculations that several graphical methods of representing the relation have been devised. In one form of representation, illustrated by Fig. 14, the per-unit impedance is expressed in terms of its real and imaginary components as $z_L = r_L + jx_L$, and loci of constant per-unit resistance (r_L) and constant per-unit reactance (x_L) are plotted in the Γ_L plane. In Fig. 14*A*, the horizontal and vertical lines form a rectangular-coordinate system for plotting the complex reflection coefficient $\Gamma_L = \gamma_r + j\gamma_i$. The solid-line circles are loci of constant r_L, and the dashed-line circles are loci of constant x_L. The real and imaginary components of Γ_L that correspond to a particular value of load impedance may therefore be read from the chart at the intersection of the proper r_L and x_L circles.

Suppose, for example, that the load impedance of the 50-ohm line of Fig. 13*A* is known to be $50 + j57.8$ ohms. The per-unit load impedance $z_L = 1.0 + j1.16$ is located at point A in Fig. 14*A* at the intersection of the $r_L = 1.0$ circle and the $x_L = 1.16$ circle. The corresponding value of Γ_L read from the chart is $0.25 + j0.433$ or, in exponential form, $0.5\epsilon^{j60°}$.

That the loci of constant per-unit resistance and reactance are the

circles drawn in Fig. 14*A* may be shown by first expressing Eq. 32 in the form

$$z_L = r_L + jx_L = \frac{1 + \gamma_r + j\gamma_i}{1 - \gamma_r - j\gamma_i}$$

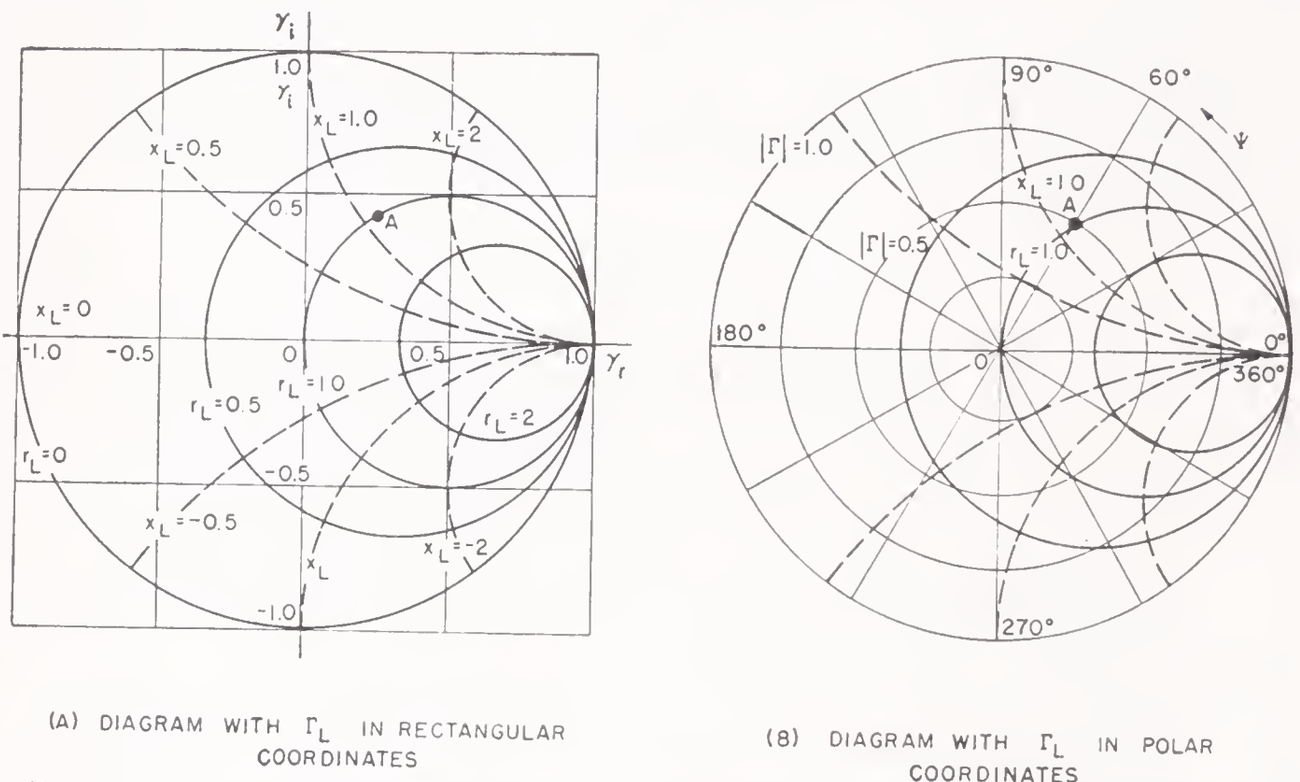
and then rationalizing the fraction to obtain

$$r_L + jx_L = \frac{1 - \gamma_r^2 - \gamma_i^2}{(1 - \gamma_r)^2 + \gamma_i^2} - \frac{j2\gamma_i}{(1 - \gamma_r)^2 + \gamma_i^2} \tag{34}$$

In order for the two complex numbers of Eq. 34 to be equal, the real components must be equal and the imaginary components must be equal. Equating the real components and completing the square of the terms involving γ_r leads to the expression

$$\left(\gamma_r - \frac{r_L}{1 + r_L}\right)^2 + \gamma_i^2 = \left(\frac{1}{1 + r_L}\right)^2 \tag{35}$$

which is the equation of a circle in the γ_r, γ_i plane with its center on the real axis at the point $r_L/(1 + r_L) + j0$ and with radius $1/(1 + r_L)$. The



(A) DIAGRAM WITH Γ_L IN RECTANGULAR COORDINATES

(B) DIAGRAM WITH Γ_L IN POLAR COORDINATES

FIG. 14. Graphical representation of load impedance in reflection-coefficient plane.

constant-resistance circles in Fig. 14*A* are constructed in accordance with Eq. 35. Observe that all resistance circles pass through the point $(1 + j0)$, as may be seen by substituting the coordinates of this point for γ_r and γ_i in Eq. 35.

Equating the imaginary components of Eq. 34 and completing the square of the terms involving γ_i yields the relation

$$(\gamma_r - 1)^2 + \left(\gamma_i - \frac{1}{x_L}\right)^2 = \frac{1}{x_L^2} \tag{36}$$

which is the equation of a circle with center at $1 + j(1/x_L)$ and radius $1/x_L$. Positive values of x_L correspond to inductive reactance and yield circles above the real axis; negative values of x_L correspond to capacitive reactance and give circles below the real axis. Because the centers of the reactance circles are displaced from the real axis by the amount of their radii, all reactance circles pass through the point $1 + j0$. Only the parts of the reactance circles that lie within the locus of $r_L = 0$ are drawn in Fig. 14*A*.

Rectangular-coordinate components of Γ_L are employed in Fig. 14*A* because of their usefulness in the derivation of the r_L and x_L circles. In actual use of the chart, however, the polar coordinates of Fig. 14*B* are more convenient. The circles centered at the origin in this figure are loci of constant $|\Gamma_L|$, and the radial lines are loci of constant ψ. At point *A*, for example, where $z_L = 1.0 + j1.16$, the magnitude and angle of Γ_L can be read directly as $\Gamma_L = 0.5\epsilon^{j60°}$.

8. Input Impedance of Transmission Lines. The input impedances of shorted-end and open-end lines are shown by Eq. 19 (p. 482) and Eq. 23 (p. 485) to depend upon the electrical length and the characteristic resistance of the line. More generally, the input impedance of a line section depends upon the electrical length, the characteristic resistance, and the terminating impedance of the line. The nature of the relationship is considered in this article.

Let the complex numbers E and I represent, as before, the voltage and current at any point P along a transmission line, the distance between P and the load being s. Let Z designate the ratio E/I; that is, let Z be the input impedance of the section of line between point P and the load. Impedance Z is then a function of s and has the special value Z_L (the load impedance) if $s = 0$ or the special value Z_d (the input impedance of the entire line of length d) if $s = d$. Also let Γ, the *reflection coefficient at point P*, designate the ratio E^-/E^+ of the reflected-wave and incident-components of E at point P. The value of Γ depends upon s, and Γ has the special value Γ_L (reflection coefficient at the load) for $s = 0$ and the special value Γ_d for $s = d$.

The relation between the impedance Z at point P and the corresponding reflection coefficient Γ is the same as the analogous load relation of Eq. 32, p. 493. In fact,

$$Z = \frac{E}{I} = \frac{E^+ + E^-}{I^+ + I^-} = \frac{E^+ + E^-}{E^+/R_c - E^-/R_c} = R_c \frac{1 + E^-/E^+}{1 - E^-/E^+} = R_c \frac{1 + \Gamma}{1 - \Gamma}$$

or

$$z = \frac{1 + \Gamma}{1 - \Gamma} \tag{37}$$

where $z = Z/R_c$ is the per-unit input impedance at point P. The relation of Γ to Γ_L may be determined from the relation of E^+ to $E_L{}^+$ and E^- to $E_L{}^-$ as follows:

$$\Gamma = \frac{E^-}{E^+} = \frac{E_L{}^- \epsilon^{-j\beta s}}{E_L{}^+ \epsilon^{j\beta s}} = \frac{E_L{}^-}{E_L{}^+} \epsilon^{-j2\beta s}$$

or

$$\Gamma = \Gamma_L \epsilon^{-j2\beta s} \tag{38}$$

The relation of Γ_L to the per-unit load impedance is specified by Eq. 33, p. 493, which equation is repeated here:

$$\Gamma_L = \frac{z_L - 1}{z_L + 1} \tag{39}$$

The input impedance at any point P can be determined for any load impedance by application of Eqs. 37 to 39. By means of Eq. 39, the load reflection coefficient is obtained from the per-unit load impedance, and from Eq. 38, the reflection coefficient Γ at the point of interest is determined. Equation 37 then yields in per-unit form the desired input impedance.

A direct relation between input impedance and load impedance is obtained if the above steps are performed in terms of letter symbols. The result is

$$z = \frac{1 + \dfrac{z_L - 1}{z_L + 1} \epsilon^{-j2\beta s}}{1 - \dfrac{z_L - 1}{z_L + 1} \epsilon^{-j2\beta s}} = \frac{z_L(1 + \epsilon^{-j2\beta s}) + (1 - \epsilon^{-j2\beta s})}{(1 + \epsilon^{-j2\beta s}) + z_L(1 - \epsilon^{-j2\beta s})}$$

If numerator and denominator are divided by $1 + \epsilon^{-j2\beta s}$, and if the quantity $(1 - \epsilon^{-j2\beta s})/(1 + \epsilon^{-j2\beta s})$ is replaced by its equal $j \tan \beta s$, the relation obtained is

$$z = \frac{z_L + j \tan \beta s}{1 + j z_L \tan \beta s} \tag{40}$$

or in terms of actual impedances instead of per-unit impedances,

$$Z = R_c \frac{Z_L + j R_c \tan \beta s}{R_c + j Z_L \tan \beta s} \tag{41}$$

Sometimes Eq. 40 or 41 furnishes a convenient and simple method of calculating input impedance, but more often the three-step process of Eqs. 37 to 39 is simpler and may be employed more advantageously. Especially is the three-step process useful if the chart of Fig. 14 is utilized to relate reflection coefficients and impedances. The point in the chart at which r_L and x_L correspond to a given load impedance is located. The reflection coefficient Γ_L corresponding to this point is changed to the

reflection coefficient $\Gamma = \Gamma_L \epsilon^{-j2\beta s}$, corresponding to a point at a distance s from the load by decreasing the angle of Γ_L by the amount $2\beta s$, that is, by rotating the representative point on the chart through an angle $2\beta s$ *clockwise* about the origin. The new point thus determined specifies Γ, and since the relation of Γ to z is the same as that of Γ_L to z_L, the per-unit input impedance may be read from the resistance and reactance circles through the new point.

As an example of this procedure, consider the line of Fig. 13*A*, and assume that the input impedance of the 75-deg section to the right of

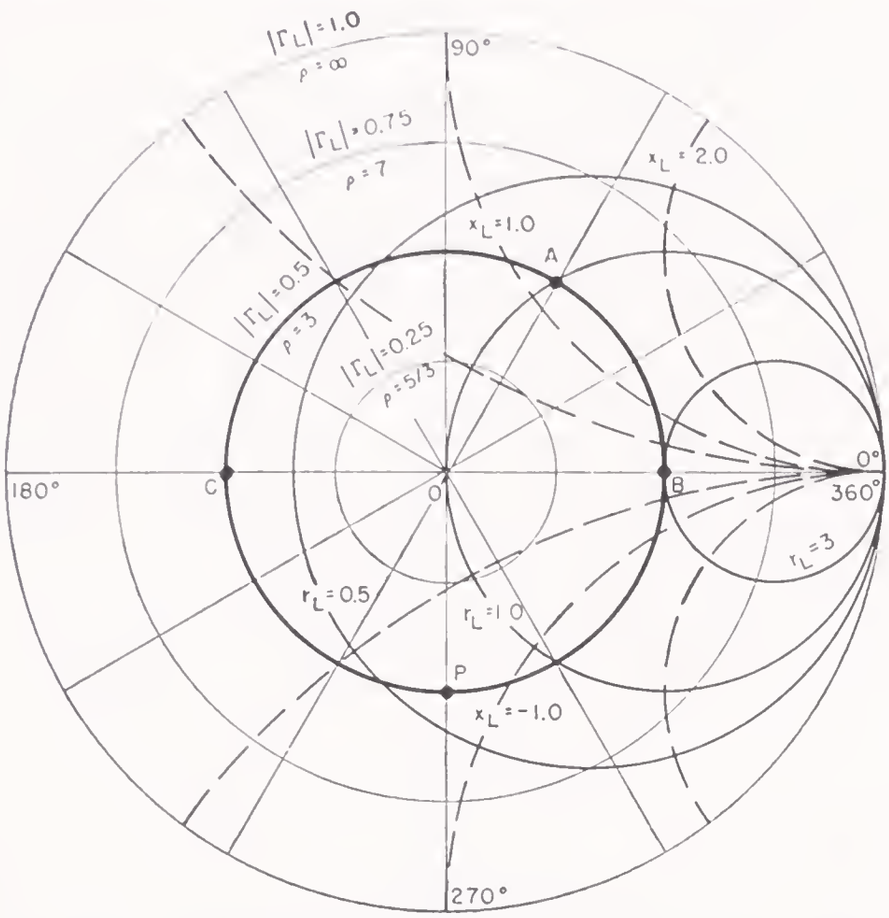


FIG. 15. Input impedance of line having load reflection coefficient of $0.5\epsilon^{j60°}$.

point P is to be found. The load reflection coefficient $0.5\epsilon^{j60°}$ is represented by point A in Fig. 15. The reflection coefficient at point P is obtained by a clockwise rotation of 2×75 deg in Fig. 15 and has the value $0.5\epsilon^{-j90°}$, as indicated by point P in Fig. 15. The corresponding input impedance read from the chart is $0.6 - j0.8$ on a per-unit basis, or $30 - j40$ ohms for a 50-ohm line.

The general way in which input impedance of a line varies as the electrical length changes may be determined either by reference to Eq. 41 or by application of the three-step process to the chart of Fig. 15. The change in electrical length may be the result of a change of physical length at constant frequency or a change in frequency (and hence in β) for a line of fixed length. As the electrical length increases from zero, the representative point on a chart such as Fig. 15 moves clockwise around a circle

of constant $|\Gamma_L|$ because the angle, but not the magnitude, of Γ depends upon electrical length. For electrical lengths that place the impedance point on the horizontal axis, the input impedance is a resistance. If the point lies above the horizontal axis, Z corresponds to the impedance of a combination of resistance and inductance; if the point is below the horizontal axis, Z comprises resistance and capacitive reactance. A complete cycle of impedance values is obtained in a 180-deg change in electrical length, and the cycle is repeated in each additional 180-deg length. The periodic nature of Z can be seen by following the motions of a representative point on the chart or by examination of Eq. 41, in which $\tan \beta s$ has a period of 180 deg.

For the example considered, the representative point starts at A in Fig. 15 and traverses the circle through points B, P, and C. For short electrical lengths, corresponding to points between A and B, the input impedance corresponds to a combination of resistance and inductance. At point B, the input impedance is a resistance of per-unit value 3, between points B and C the impedance is that of a resistance-capacitance combination, and at C the impedance is a resistance of per-unit value $\frac{1}{3}$. Between points C and A, the impedance is again that of an R-L combination, and for a 180-deg electrical length the representative point is again at A and the input impedance is the same as the load impedance.

The circles of constant $|\Gamma_L|$ in Fig. 15 are also circles of constant standing-wave ratio ρ because of the relation between ρ and $|\Gamma_L|$ specified by Eq. 30, p. 490. The value of ρ corresponding to each value of $|\Gamma_L|$ is indicated in Fig. 15. Another important relationship exists between ρ and the maximum and minimum values of the input impedance obtained as the line length is varied. The input impedance is maximum at a point such as B in Fig. 15, where the reflection coefficient is a positive real number and the standing-wave patterns of voltage and current have, respectively, maximum and minimum values. At such a point, the total voltage and total current are in phase, so that the maximum impedance is a resistance and has the value

$$\frac{|E|_{\max}}{|I|_{\min}} = \frac{|E|_{\max}}{|I^+| - |I^-|} = R_c \frac{|E|_{\max}}{|E^+| - |E^-|} = R_c \frac{|E|_{\max}}{|E|_{\min}} = R_c \rho$$

or the *maximum impedance is a resistance of per-unit value* ρ. Similarly, the *minimum impedance* (which occurs at a point such as C in Fig. 15, where the reflection coefficient is a negative real number) *is a resistance of per-unit value* $1/\rho$.

Of particular importance are the input impedances of lines having certain special values of electrical length or load impedance. Two such special cases, lines having short-circuit and open-circuit terminations, are discussed in Arts. 3 and 4. Note that the value of Z specified by Eq. 41

simplifies to $jR_c \tan \beta s$ when $Z_L = 0$, in accordance with Eq. 19 (p. 482) for a shorted-end line. Similarly, when Z_L is infinite, Eq. 41 yields the value $-jR_c \cot \beta s$ for Z in agreement with Eq. 23 (p. 485) for an open-end line.

A third special case is the half-wave line, or line having an electrical length of 180 deg. Because of the periodic property of the input impedance, *the input impedance of a half-wave line is equal to the load impedance.* The load voltage is of the same magnitude as the input voltage, but of opposite polarity. Thus a half-wave line may be thought of as a 1:1 transformer connected so as to provide a polarity reversal.

A fourth special case is the quarter-wave line or line of 90 deg electrical length. *On a per-unit basis, the input impedance of a quarter-wave line is the reciprocal of the load impedance.* As proof of this statement, consider Eq. 37 for per-unit input impedance, and note from Eq. 38 that Γ for a quarter-wave line equals $\Gamma_L \epsilon^{-j180^\circ} = -\Gamma_L$. The input impedance of such a line is therefore

$$z = \frac{1 + \Gamma}{1 - \Gamma} = \frac{1 - \Gamma_L}{1 + \Gamma_L}$$

According to Eq. 32 (p. 493),

$$z_L = \frac{1 + \Gamma_L}{1 - \Gamma_L}$$

and thus $z = 1/z_L$. In terms of actual impedance, $z = Z/R_c$ and $z_L = Z_L/R_c$. Hence,

$$Z = \frac{R_c^2}{Z_L} \tag{42}$$

Equation 42 indicates that a quarter-wave line transforms an impedance of large magnitude into one of small magnitude, and vice versa. This reciprocal relation between input and load impedance makes a quarter-wave line useful as an impedance-changing device (Art. 15). In such an application, the line section is called a quarter-wave transformer.

Because input impedance is not altered by the addition of multiple half-wave sections of line, special impedance properties are associated not only with half-wave and quarter-wave lines but also with lines comprising any integral number of quarter-wave sections. If the number of quarter-wave sections is even, the input impedance equals the load impedance, as for a half-wave line. If the number of quarter-wave sections is odd, the per-unit input impedance is the reciprocal of the per-unit load impedance, as for a quarter-wave line.

9. Transmission-line Charts. Several graphical representations of the relation between reflection coefficient and per-unit impedance are widely employed in the manner described in Art. 8, both for the quantita-

tive calculation of impedance values in transmission-line systems and for the qualitative determination of the nature of impedance changes that accompany changes in line length or terminal impedance. Of particular importance is the *Smith chart,*[1] one form of which is illustrated in Fig. 16 and differs from the Γ-plane diagram of Fig. 14 only in minor respects.

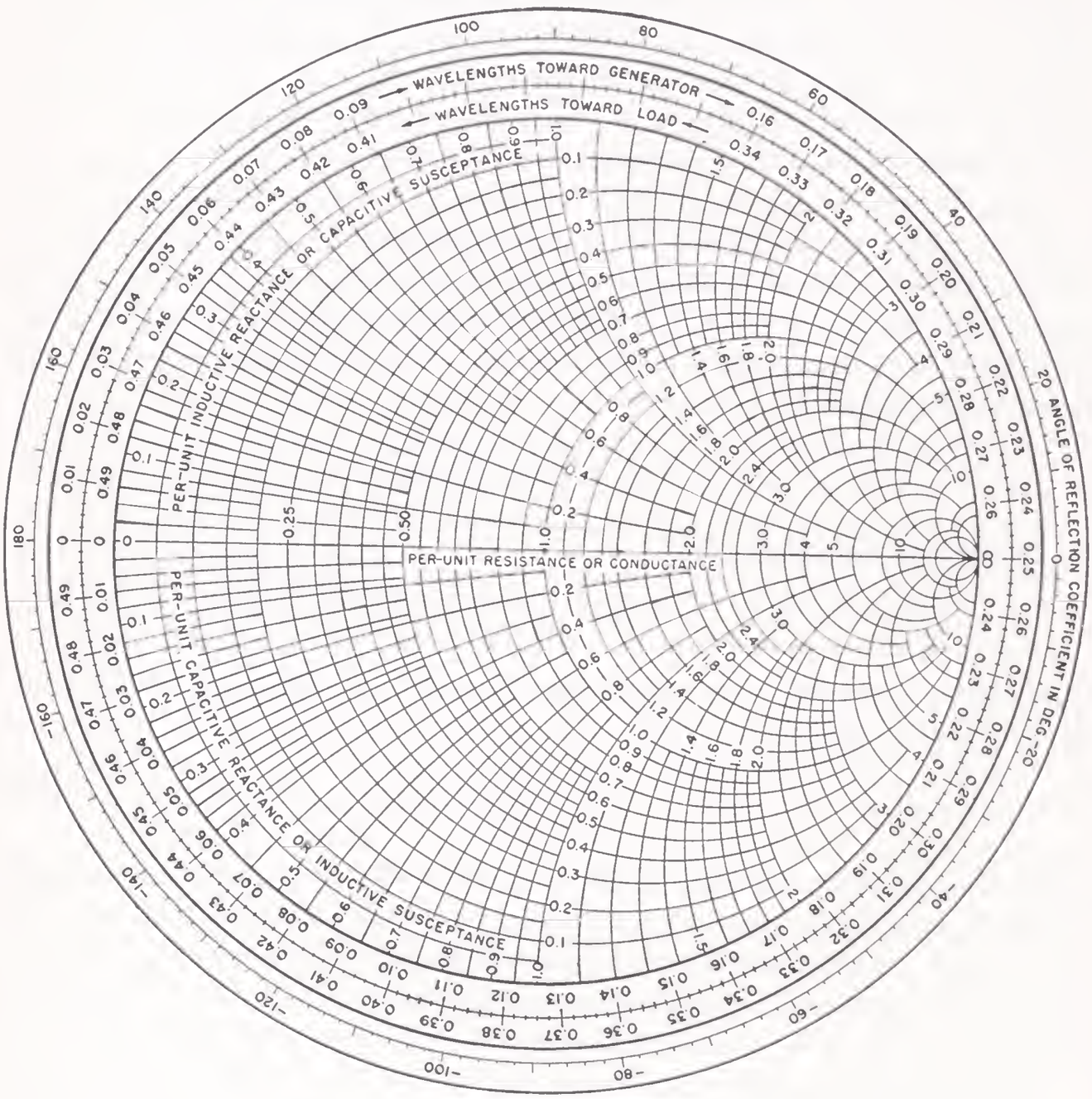

WAVELENGTHS TOWARD GENERATOR
WAVELENGTHS TOWARD LOAD
PER-UNIT INDUCTIVE REACTANCE OR CAPACITIVE SUSCEPTANCE
PER-UNIT RESISTANCE OR CONDUCTANCE
PER-UNIT CAPACITIVE REACTANCE OR INDUCTIVE SUSCEPTANCE
ANGLE OF REFLECTION COEFFICIENT IN DEG.

FIG. 16. The Smith chart.

In Fig. 16 many constant-resistance and constant-reactance circles are drawn to permit reasonable accuracy in reading impedance values. To avoid confusion of the diagram, polar-coordinate lines for Γ are omitted. For the solution of a particular example, as many of these lines as are needed may be added with the aid of a straightedge and a compass. The scale of angle for Γ around the outer edge of the chart is helpful in locating radial coordinate lines. For convenience in measuring the

[1] Figure 16 is adapted from P. H. SMITH, "An Improved Transmission-line Calculator," *Electronics,* **17** (January, 1944), 130–133, 318, with permission.

angular motion of a representative point on the chart, scales of wavelengths toward the load and wavelengths toward the generator are included inside the angle scale for Γ in Fig. 16. Note that a half-wave change, which corresponds to an 180-deg change in βs or a 360-deg change in the angle of Γ, corresponds to a complete rotation about the chart. The location of the zero points of the wavelength scales is not important, because the scales are intended only to indicate the *change* of position in wavelengths associated with some motion of a representative point. In Fig. 16, the zero points are arbitrarily located at a point of minimum impedance.

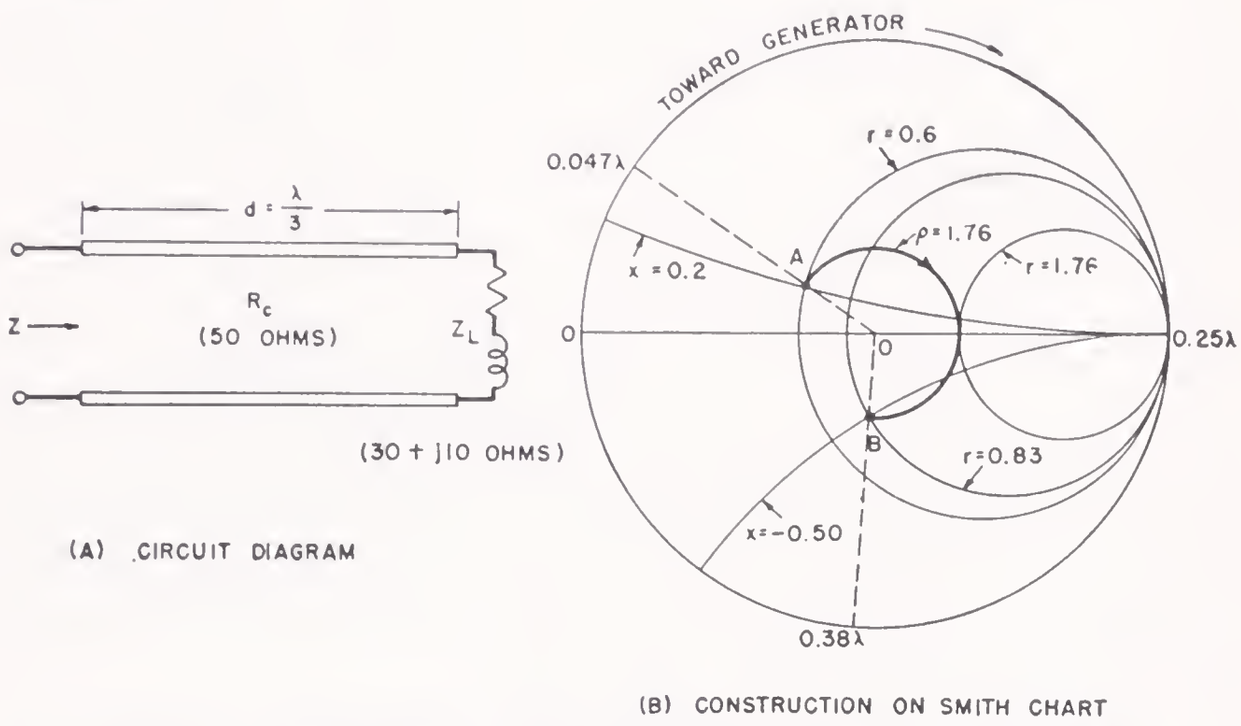


FIG. 17. Determination of input impedance by means of Smith chart.

The use of the Smith chart may be illustrated by the example of Fig. 17*A*, in which the input impedance of the 50-ohm one-third-wavelength line ($\beta s = 120$ deg) terminated by an impedance of $30 + j10$ ohms is to be found. The per-unit load impedance

$$z_L = \frac{30 + j10}{50} = 0.6 + j0.2$$

is first located at point A in Fig. 17*B*, and the corresponding reference wavelength is read on the wavelengths-toward-generator scale as 0.047λ. Because the line length is 0.333λ, the input impedance is read at the intersection of the ρ circle through A with a radial line through the point $0.047 + 0.333 = 0.38$ on the wavelength scale. This intersection is at point B, and the input impedance is therefore $0.83 - j0.50$ per unit, or $41.5 - j25.0$ ohms. The standing-wave ratio for the line is 1.76 because the ρ circle is tangent to the $r = 1.76$ circle.

The Smith chart may also be used to determine load impedance when

the standing-wave pattern is known. As an example, suppose that the standing-wave pattern has a minimum at a point 0.2λ from the load and a standing-wave ratio of 2.5, as indicated in Fig. 18*A*. A $\rho = 2.5$ circle is drawn on the chart (Fig. 18*B*), and the minimum-voltage point is located

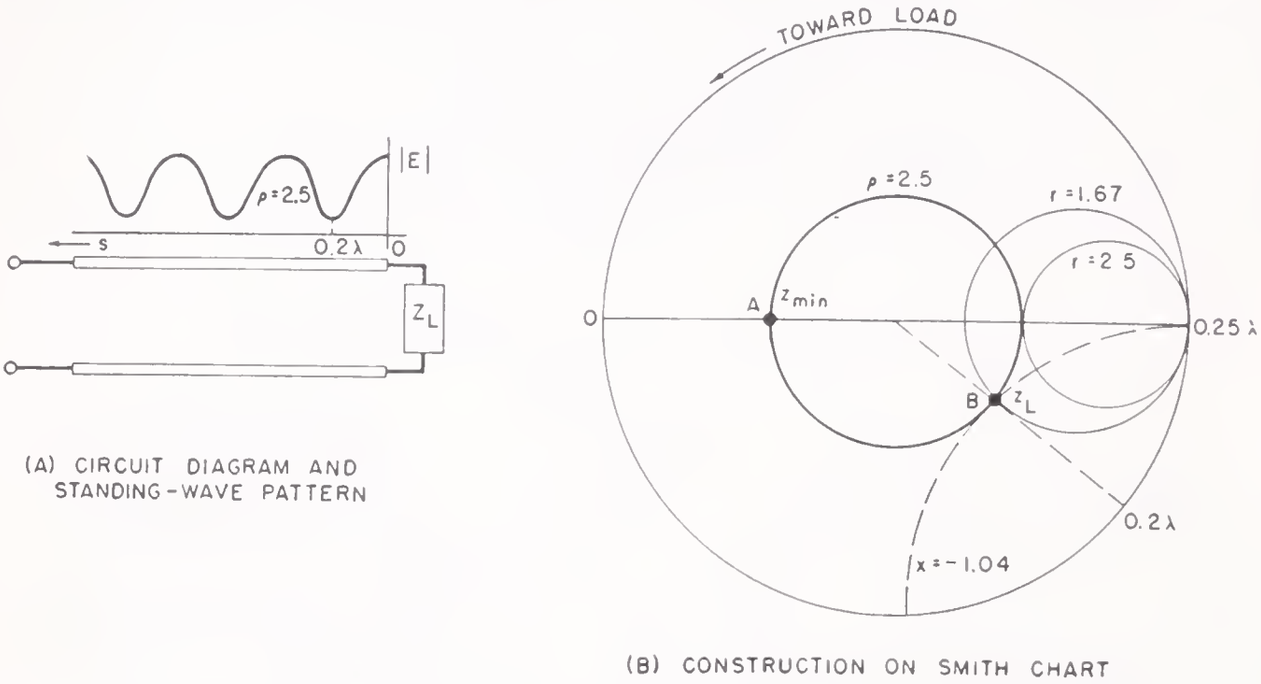


FIG. 18. Determination of load impedance by means of Smith chart.

at *A*. The load impedance is determined from point *B*, located on the $\rho = 2.5$ circle at position corresponding to a 0.2λ counterclockwise rotation from *A*. This position is located by reference to the 0.2 mark on the wavelengths-toward-load scale. At point *B*, the load impedance is $1.67 - j1.04$ per unit, or $83.2 - j52.0$ ohms.

The Smith chart is a graphical representation of the relation between Γ and z obtained by plotting specific values of z (points on the resistance and reactance circles) in the Γ plane. An alternate graphical representation of the relation is obtained by plotting values of Γ in the z plane. Customary practice is to plot loci of constant magnitude and constant angle of Γ in the z plane. These loci are circles, as indicated in Fig. 19, and the chart thus obtained is called a *z-plane circle diagram.*

The z-plane chart may be thought of as being obtained by a distortion of the Smith chart in which the region near the right-hand end of the horizontal axis in Fig. 16 is expanded tremendously and other portions of the chart are distorted to a smaller degree. The expansion is continued until the constant-resistance circles of the Smith chart become the vertical-coordinate lines in Fig. 19 and the constant-reactance circles become the horizontal-coordinate lines. In this transformation, the circles of constant magnitude of reflection coefficient (circles centered at the origin in the Smith chart) become the circles surrounding $1 + j0$ in Fig. 19. The radial lines of constant angle of the reflection coefficient in the Smith chart become the arcs of circles through the point $1 + j0$ in

Fig. 19. Note that a z-plane diagram corresponding to the entire Smith chart would have infinite extent vertically and to the right, and only the portion of such a chart applicable for standing-wave ratios of not more than 5 is drawn in Fig. 19.

The scales for per-unit resistance and per-unit reactance are marked on the horizontal and vertical axes of Fig. 19. The $|\Gamma|$ values for which

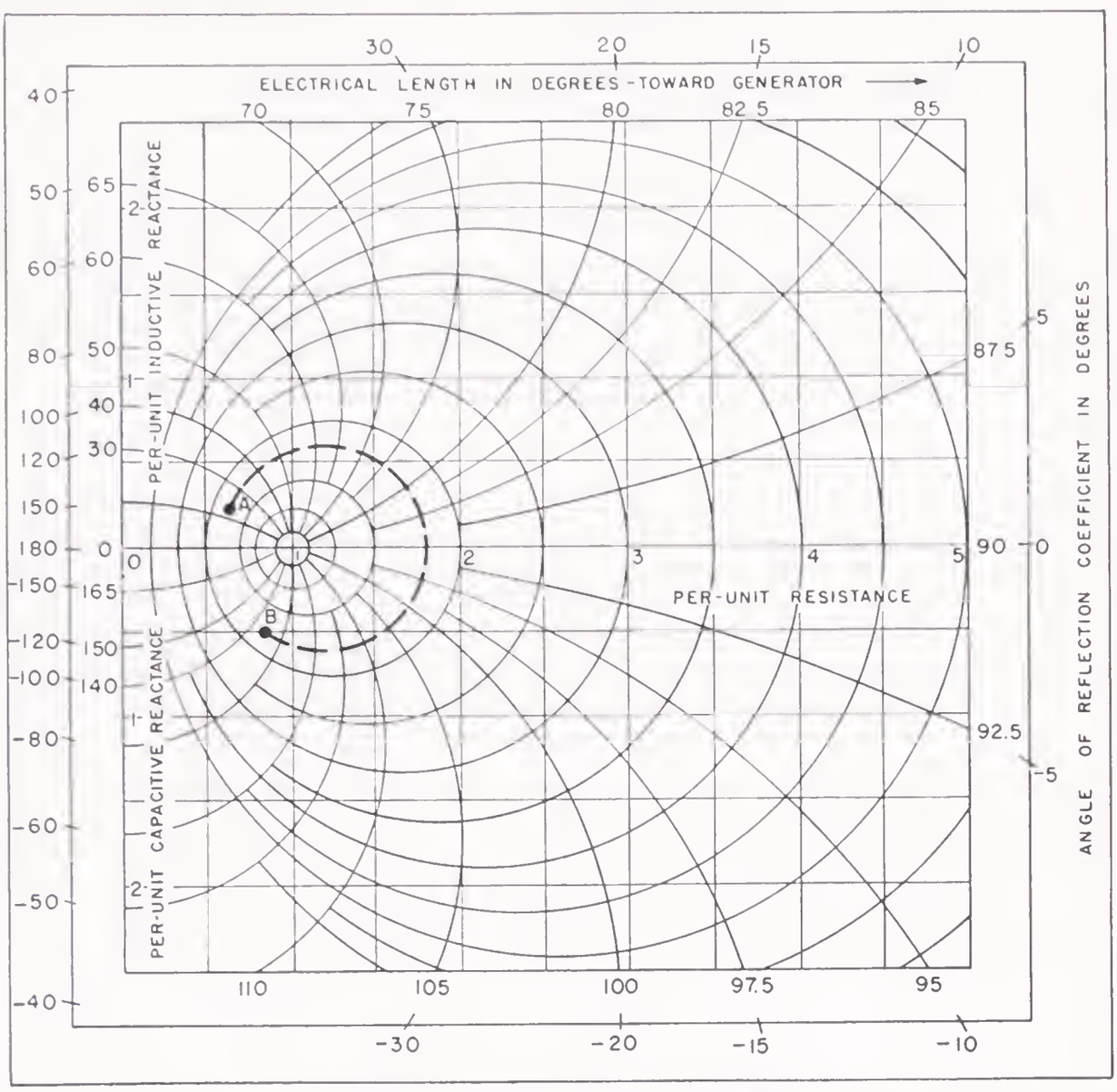


FIG. 19. The z-plane circle diagram.

circles are drawn are not indicated, because these circles are also circles of constant standing-wave ratio. The ρ values can be read from the per-unit resistance scale at the right-hand intersections of the circles with the horizontal axis, and the $|\Gamma|$ values, if needed, can be calculated from Eq. 30, p. 490. The scale for the angle of Γ is marked around the outside of the chart, and a scale of βs measured from a point of minimum voltage is included inside the Γ-angle scale. This scale is used in the same manner as are the wavelength scales of the Smith chart.

As an example of the use of the z-plane circle diagram, consider again the problem of finding the input impedance of the line section in Fig. 17*A*.

The per-unit load impedance $z_L = 0.6 + j0.2$ is represented by point A in Fig. 19. This point is on the $\rho = 1.76$ circle at an angle of 17 deg on the electrical-length scale. Therefore, the input impedance is represented by point B located at the intersection of the same ρ circle with the angle of circle $17 + 120 = 137$ deg. The input impedance is $0.83 - j0.50$ per unit in agreement with the value obtained from the Smith chart.

That the loci of constant magnitudes of Γ are the circles surrounding $1 + j0$ in Fig. 19 may be proved in the following manner. By Eq. 37 (p. 495), $\Gamma = (z - 1)/(z + 1)$, and thus if z is written as $r + jx$, the square of the magnitude of Γ is

$$|\Gamma|^2 = \frac{(r - 1)^2 + x^2}{(r + 1)^2 + x^2} \tag{43}$$

By rearranging this equation in powers of r and then completing the square in the terms involving r, the following expression may be obtained:

$$\left(r - \frac{1 + |\Gamma|^2}{1 - |\Gamma|^2}\right)^2 + x^2 = \left(\frac{2|\Gamma|}{1 - |\Gamma|^2}\right)^2 \tag{44}$$

Equation 44 is the equation of a circle in the z plane with its center on the real axis at $(1 + |\Gamma|^2)/(1 - |\Gamma|^2) + j0$ and with radius $2|\Gamma|/(1 + |\Gamma|^2)$.

Similarly, the loci of constant angles of Γ may be shown to be the arcs of circles through $1 + j0$ in Fig. 19. Equating the angles of the complex-number equation $\Gamma = (z - 1)/(z + 1)$ yields the expression

$$\psi = \alpha - \beta \tag{45}$$

where ψ is the angle of Γ, α is the angle of $(r - 1) + jx$, and β is the angle of $(r + 1) + jx$. Thus

$$\tan \alpha = \frac{x}{r - 1} \qquad \text{and} \qquad \tan \beta = \frac{x}{r + 1}$$

and

$$\tan \psi = \tan (\alpha - \beta) = \frac{\tan \alpha - \tan \beta}{1 + \tan \alpha \tan \beta} = \frac{\dfrac{x}{r - 1} - \dfrac{x}{r + 1}}{1 + \dfrac{x^2}{r^2 - 1}}$$

$$= \frac{2x}{r^2 + x^2 - 1}$$

This expression may be rearranged in powers of x and the square in these terms completed, with the result

$$r^2 + (x - \cot \psi)^2 = \csc^2 \psi \tag{46}$$

The loci of constant ψ are therefore circles with centers on the imaginary axis, the locations of the centers being the points $0 + j \cot \psi$ and the radii being $\csc \psi$.

Both the Smith chart and the z-plane circle diagram provide convenient means of determining impedance relations on a lossless transmission line. An advantage of the z-plane diagram is that coordinate lines for both Γ and z appear on the chart, so that no additional construction lines need be drawn to solve specific problems. This advantage is more than offset, however, by the fact that all values of r and x determine points within the $|\Gamma| = 1$ circle of the Smith chart whereas the z-plane diagram is of infinite extent. In order to contain points for high standing-wave ratio, the z-plane diagram must be drawn to such a small scale that coordinates of points of low standing-wave ratio cannot be read with accuracy. Thus the Smith chart is the diagram most commonly used for transmission-line calculations. It is employed in the remainder of this book not only in connection with transmission lines, but also for the determination of impedance relations in waveguides, as explained in Chap. VIII.

10. Input Admittance. Frequently a load impedance is shunted across a transmission line at a point remote from the line terminations, or the output terminals of one line are joined in a parallel connection to the input terminals of two or more other lines. Because *admittances add at a parallel junction*, calculations on such systems are considerably simplified if the input admittances, rather than the input impedances, of the line sections are employed. Input admittance can be determined by first calculating the input impedance by means of any of the equations or charts of the preceding articles and then equating admittance to the reciprocal of impedance. More direct ways are available, however, for using the Smith chart and the z-plane diagram with admittances, and these direct procedures are very convenient for use with parallel systems.

The input admittance to any section of line is $Y = 1/Z$, where Z is the input impedance. The per-unit input admittance $y = Y/G_c$ (where $G_c = 1/R_c$ is the characteristic conductance of the line) is therefore

$$y = \frac{Y}{G_c} = \frac{1/Z}{1/R_c} = \frac{R_c}{Z} = \frac{1}{z} \tag{47}$$

The Smith chart is based upon the relation $z = (1 + \Gamma)/(1 - \Gamma)$. Because $y = 1/z$, this relation may be written

$$y = \frac{1 - \Gamma}{1 + \Gamma} = \frac{1 + (-\Gamma)}{1 - (-\Gamma)} \tag{48}$$

Thus the relation of y to $-\Gamma$ is exactly the same as the relation of z to $+\Gamma$, and the Smith chart may be considered a plot of values of y in the $-\Gamma$ plane. As indicated by the scale designations in Fig. 16 (p. 500), the resistance circles become circles of constant per-unit conductance and the reactance circles become circles of constant per-unit susceptance. The change from the Γ to the $-\Gamma$ plane can be accounted for by rotating

the scale for angle of the reflection coefficient 180 deg around the chart. Note that the change from the Γ to the $-\Gamma$ plane has no significance in the calculation of the admittance at one point of the line from that at another point. The rotation of the scale of angle of Γ is of interest only when the value of Γ is given or is to be found.

As an example of the use of the Smith chart with admittances, let it be required to find the input admittance of the shorted-end line in Fig. 20. The short circuit constitutes a load of infinite admittance, so that the representative point for the load is located at the right-hand end of the horizontal axis of the Smith chart (p. 500). This point is at 0.25 on the wavelength scale, and thus the representative point at the input is located on the outside ($\rho = \infty$ or $g = 0$) circle at a reading on the wavelengths-toward-generator scale of $(0.25 + 0.333) - 0.5 = 0.083$, corresponding to a clockwise rotation of one-third wavelength from the initial point. The input admittance is therefore $0 + j0.58$ per unit, or for a 0.01-mho (100-ohm) line, $0 + j0.0058$ mho. Observe that, if a reflection coefficient is desired, its angle must be read from the angle scale rotated 180 deg. For example, the reflection coefficient at the load is $1\epsilon^{j180^\circ} = -1$, as it must be at a short circuit.

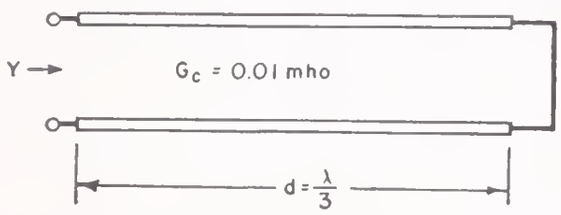


FIG. 20. Line employed as example in input-admittance calculation.

Frequently in transmission-line problems, a change from per-unit impedance to per-unit admittance (or vice versa) is required. This change requires that the reciprocal of a complex number be found, and a procedure that is shorter than a slide-rule calculation of the reciprocal is to rotate the point representing the given per-unit value 180 deg either way around a circle of constant ρ. This procedure is correct because the input per-unit impedance of a quarter-wave line is the reciprocal of the per-unit load impedance (see Eq. 42, p. 499). The input impedance of the line in Fig. 20, for example, is obtained by an 180-deg rotation along the $\rho = \infty$ circle from the $0 + j0.58$ per-unit-admittance point, and the impedance is $0 + j1.73$ per unit, or $0 + j173$ ohms.

Because the relation of $-\Gamma$ to y is the same as that of Γ to z (see Eq. 48), the z-plane diagram of Fig. 19, p. 503, may be considered alternatively as a plot of values of $-\Gamma$ in the per-unit admittance plane. Thus the z-plane diagram, like the Smith chart, may be used equally well with per-unit admittances, but the scale of angle for Γ must be shifted 180 deg if values of Γ corresponding to admittance values are to be determined.

11. Transient and Steady-state Operation of Transmission Lines. In the preceding articles, the generator that supplies power to a transmission line is assumed to be a fixed sine-wave voltage source in series

with a resistance equal to the characteristic resistance of the line. With such a generator, any wave reflected from the load causes no further reflections when it reaches the generator, and (if the load is purely resistive) steady-state operation is reached after one round trip of the voltage and current waves.

Also of interest are generators having values of internal impedance different from R_c and generators (such as triode oscillators, klystrons, and magnetrons—see Chaps. X and XI) that cannot properly be represented as a fixed voltage source in series with a linear impedance. When such generators are suddenly connected to a line, the transient obtained is more complicated and of longer duration than the one obtained with a generator of internal resistance R_c. A wave originating at the source and reflected at the load suffers a second reflection when it returns to the source. The new forward-going wave is also reflected at the load, and the process repeats, the wavefront traveling back and forth many times and growing smaller at each reflection. Before steady-state operation can be attained, the traveling wavefront must be reduced to negligible proportions. If the impedances at the ends of the line are not simple resistances, the transient may be further complicated by the growth of current in terminal inductors and growth of charge on terminal capacitors during the transient period.

However involved the transient process may be, the final steady-state condition, which alone is of interest in this chapter, can be represented by a series of sine waves (often with a variety of different phases and amplitudes) traveling in the direction from source to load and a second series of waves traveling in the reverse direction. Because all waves travel with the same speed and are of the same frequency, all the forward-going waves can be summed and considered as a single resultant forward-going sine wave. Similarly, all the returning waves can be added, and the sum considered to be a single return sine wave. Thus the steady-state distribution of voltage and current on line can always be thought of as the sum of *two* traveling waves, one proceeding in each direction along the line.

Observe that, although the total forward-going wave is different from the initial wave from the source and the total returning wave is different from the first reflected wave, *the relationship of the total return wave to the total forward wave is determined by the load* in exactly the same way as is the relationship of any component wave reflected from the load to the corresponding incident wave. The total return wave must have such a value relative to the total forward wave that the resultant voltage at the load is the product of the resultant load current by the load impedance. Thus the voltage at the load of the total return wave is Γ_L times the corresponding voltage of the total forward wave, where Γ_L is the load

reflection coefficient defined in Art. 6. Similarly, the reflection coefficient Γ specifies the ratio of return-wave voltage to forward-wave voltage at any point of the line, and thus *the standing-wave ratio and the position of the standing-wave pattern are determined by the load and are independent of the characteristics of the generator* (provided the generator supplies sinusoidal waves of voltage and current).

Although the generator does not affect the shape or position of the standing-wave pattern, it does, of course, determine the over-all scale of magnitude for the pattern. An increase in power delivered by the generator increases the voltage and current at every point in the standing-wave pattern. In order to determine the voltage and current magnitudes, the input impedance of the line is first determined (by use of the Smith chart, for example) from given values of load impedance and line length. This input impedance is the load impedance for the generator and may be used to determine the power, voltage, or current delivered by the generator. For a linear generator, this determination involves a calculation upon a series circuit comprising the source voltage and internal impedance and the input impedance of the line. For a nonlinear source, such as a vacuum-tube oscillator, performance charts (for example, the Rieke diagrams discussed in Art. 10 of Chap. XI) specifying the power and frequency of the oscillator as functions of the load impedance are used.

As an example of the relation of the component waves in the transient period to the total forward-going and return waves of the steady-state condition, consider the illustrative circuit of Fig. 21*A*. To simplify the calculations, the terminating impedance is assumed to be a resistance (of per-unit value 1.5) and the line is assumed to be exactly one wavelength long. Switch S is assumed to be closed at $t = 0$, the instant at which e_g is rising through its zero value.

When switch S is first closed, the input impedance of the line equals the characteristic resistance, 100 ohms. Thus the initial wave has a peak value of

$$100 \times \frac{100}{100 + 150} = 40 \text{ volts}$$

This initial wave requires one period of duration T to traverse the one-wavelength line. Therefore, at the instant $t = T$, the voltage distribution along the line is represented by the traveling wave e_1^+ in Fig. 21*B*.

The reflection coefficient at the load, corresponding to a per-unit impedance of $1.5 + j0$, is 0.2, and therefore the wave e_1^- reflected from the load has a peak value of $40 \times 0.2 = 8$ volts. At the instant $t = 2T$, this reflected wave has traversed the entire line and added the component labeled e_1^- to the voltage on the line. The incident wave, having

progressed exactly 1 cycle between the instants $t = T$ and $t = 2T$, is again as shown by the e_1^+ wave in Fig. 21*B*; and thus the total voltage at the time $t = 2T$ is the sum of the e_1^+ and e_1^- waves. The reflection coefficient at the generator is also 0.2, and thus the e_2^+ wave of Fig. 21*B* is obtained. The total voltage at time $t = 3T$ is the sum of the three component waves in this figure.

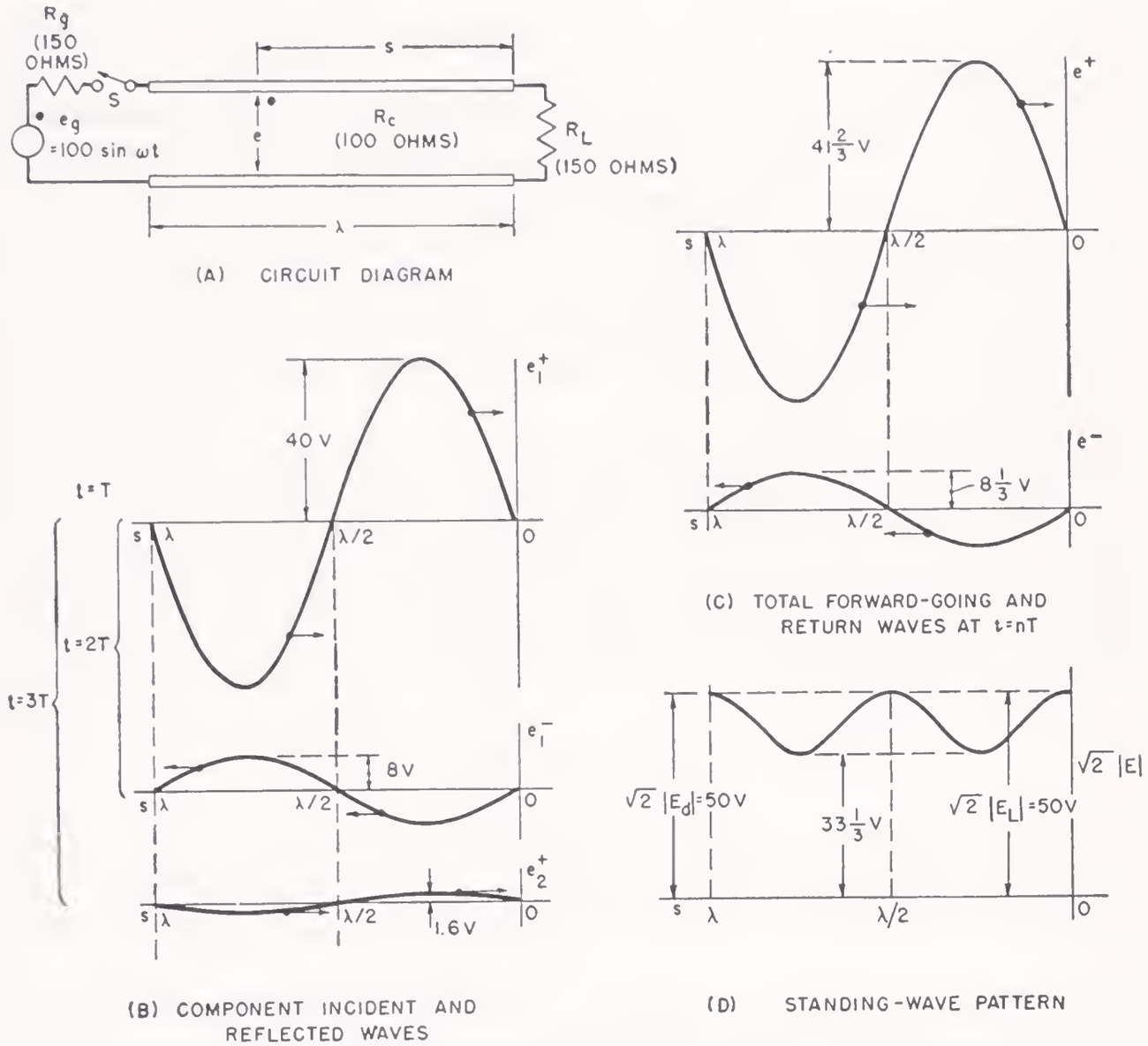


FIG. 21. Example of transient and steady-state voltage waves.

Additional traveling-wave components caused by succeeding reflections are of very small amplitude; for example, the wave caused by reflection of e_2^+ has a peak value of only 0.32 volt. Thus the total forward-going wave in the steady-state condition has an amplitude only slightly more than 41.6 volts, the amplitude of the sum of the e_1^+ and e_2^+ waves, and the total return wave has an amplitude only slightly larger than the value 8 volts obtained from the e_1^- wave. The total forward-going and return waves are indicated as e^+ and e^-, respectively, in Fig. 21*C* for an instant $t = nT$, where n is any large integer. The peak value 41⅔ volts for e^+ is obtained by adding all the forward-going waves

and is

$$\text{Peak value of } e^+ = 40 + 0.2^2 \times 40 + 0.2^4 \times 40 + \cdots$$
$$= 40\left(1 + \frac{1}{25} + \frac{1}{25^2} + \cdots\right)$$
$$= \frac{40}{1 - 1/25} = 41\frac{2}{3} \text{ volts}$$

The peak value $8\frac{1}{3}$ volts for e^- is obtained similarly from the series of return waves.

The combination of the e^+ and e^- waves results in a standing-wave pattern represented by the plot of $\sqrt{2}\,|E|$, the peak value of the total voltage in Fig. 21*D*. The forward and return waves are in phase at the generator, at the load, and at the center of the line and are 180 deg out of phase at points a quarter wave from the ends of the line. Voltage $\sqrt{2}\,|E|$ has a maximum value of 50 volts (the sum of $41\frac{2}{3}$ and $8\frac{1}{3}$ volts) at the in-phase points and a minimum value equal to the difference $33\frac{1}{3}$ volts at the quarter-wave points.

The above calculations are made only for the purpose of illustrating the relation between transient and steady-state conditions. The final results of Fig. 21*D* are much more easily obtained by steady-state calculations. Thus, in the steady state, the input impedance of the one-wavelength line is equal to the load impedance, 150 ohms. The total voltage at the input of the line therefore has the peak magnitude

$$\sqrt{2}\,|E_d| = 100 \times \frac{150}{150 + 150} = 50 \text{ volts}$$

and the load voltage has this same magnitude, because the line is one wave long. The standing-wave ratio is equal to the per-unit load impedance 1.5, and thus the minimum voltage in the standing-wave pattern is $50/1.5 = 33\frac{1}{3}$ volts.

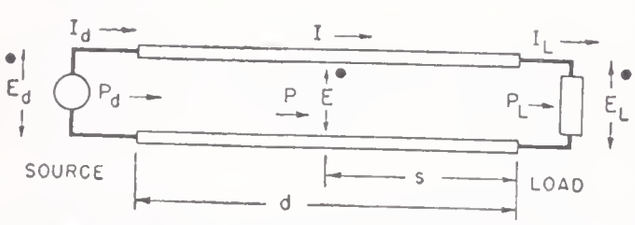


FIG. 22. Voltage, current, and power in a transmission-line system.

12. Power and Power Capacity. Of great importance in the application of transmission lines is the amount of power conveyed by a line from a source to a load. In Fig. 22, the notation for power is illustrated by reference to a line connecting a source and a load. The symbol P stands for the average power that traverses any cross section of the line in the direction of the load, P_d is the power delivered by the source, and P_L is the power consumed by the load. Voltage and current at any point of the line are represented by complex numbers, as before.

The average load power may be calculated in the same way as the power absorbed by any impedance supplied with sinusoidal waves of

voltage and current. Thus the load power is

$$P_L = |E_L|\,|I_L| \cos \theta_L$$

where $|E_L|$ and $|I_L|$ represent rms values of load voltage and current and θ_L is the phase angle between the waves of voltage and current. Similarly, with analogous definitions of symbols, the power delivered by the source to the input impedance of the line is

$$P_d = |E_d|\,|I_d| \cos \theta_d$$

and in general, the power delivered at any point of the line to the input impedance of the remaining section of line is

$$P = |E|\,|I| \cos \theta \tag{49}$$

On a lossless line, despite wide variations of $|E|$, $|I|$, and θ that may occur from point to point along the line, the product $|E|\,|I| \cos \theta$ must be constant and the value of P must be the same at all points of the line and equal to P_d and P_L.

Frequently separation of the power P into components associated with the forward-going and the returning waves is desirable. The power conveyed from the source to the load by the forward-going wave is $P^+ = |E^+|\,|I^+|$, the power factor being omitted as E^+ and I^+ are in phase and $\cos \theta^+ = 1$. Similarly the power sent back to the source by the returning wave is $P^- = |E^-|\,|I^-|$, and for a lossless line the net power transmitted from source to load is

$$P = P^+ - P^- \tag{50}$$

This equation has a practical application in connection with the directional coupler (see Art. 7, Chap. XII), a device that permits independent measurement of the power associated with the forward-going and returning waves.

Note that the correctness of Eq. 50 is *not* obvious. Linear superposition may be applied to voltages and currents on transmission lines but not, in general, to power. For example, the power of a 100-volt (rms value) wave applied to a 100-ohm line is 100 watts, and the power of a 50-volt wave on the same line is 25 watts. If two such voltage waves of the same frequency travel in phase along the same line, the total power is not the sum of the individual powers (125 watts) but rather is the power of a 150-volt wave, which is 225 watts. That superposition may be applied to power for two waves traveling in opposite direction must be regarded as a special circumstance requiring specific proof. As a matter of fact, Eq. 50 can be applied correctly only to lines for which the voltage and current of a traveling wave are in phase. Though of little concern in radar, high-loss lines having a characteristic-

impedance angle different from zero exist, and for these the power transmitted cannot be separated into components associated with the forward-going and returning waves.

Proof of Eq. 50 may be made by reference to Fig. 23. Complex numbers E and I represent voltage and current at any point of a transmission line having a resistive characteristic impedance. The voltage is represented in terms of the forward-going and return waves as $E^+ + E^-$, and the current as $I^+ + I^-$. For the definitions of angles indicated in the figure

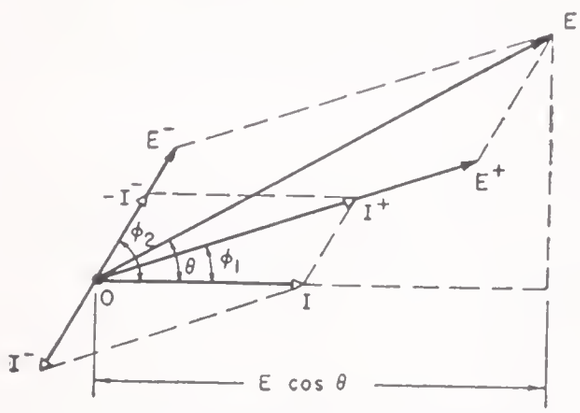


FIG. 23. Complex-number diagram used in proof of Eq. 50.

$$|I| = |I^+| \cos \phi_1 - |I^-| \cos \phi_2$$
$$|E| \cos \theta = |E^+| \cos \phi_1 + |E^-| \cos \phi_2$$

and by Eq. 49 the power transmitted is

$$\begin{aligned} P &= |E|\,|I| \cos \theta \\ &= |E^+|\,|I^+| \cos^2 \phi_1 - |E^-|\,|I^-| \cos^2 \phi_2 \\ &\qquad + (|E^-|\,|I^+| - |E^+|\,|I^-|) \cos \phi_1 \cos \phi_2 \end{aligned} \tag{51}$$

The last term in Eq. 51 is zero, because $|E^-|\,|I^+|$ and $|E^+|\,|I^-|$ are each equal to $R_c|I^+|\,|I^-|$. Figure 23 shows also that

$$|I^+| \sin \phi_1 = |I^-| \sin \phi_2$$

Squaring this relation, multiplying the result by R_c, and transposing yield

$$0 = |E^+|\,|I^+| \sin^2 \phi_1 - |E^-|\,|I^-| \sin^2 \phi_2 \tag{52}$$

and the sum of Eqs. 51 and 52 is

$$P = |E^+|\,|I^+| - |E^-|\,|I^-| = P^+ + P^-$$

A third relation for power transmitted (which is valid only for lossless lines) is

$$P = \frac{|E|_{\max}|E|_{\min}}{R_c} \tag{53}$$

Proof of Eq. 53 depends upon the fact that at a point of maximum voltage on a lossless line the total voltage and current are in phase, and thus the power transmitted is

$$P = |E|_{\max}|I|_{\min}$$

This equation is equivalent to Eq. 53, because

$$|I|_{\min} = |I^+| - |I^-| = \frac{|E^+| - |E^-|}{R_c} = \frac{|E|_{\min}}{R_c}$$

the last step being correct because $|E^+|$ and $|E^-|$ do not vary from point to point along a lossless line.

As an example of the application of the power equation, consider a line terminated in an open circuit, a short circuit, or a reactance, so that the standing-wave ratio is infinite. The power consumed by the load is zero, and therefore if the line is lossless, the power delivered by the source and the transmitted power at each point of the line must be zero. The phase angle between voltage and current in a standing-wave pattern corresponding to complete reflection is 90 deg (see Arts. 3 and 4), and thus $\cos\theta$ in Eq. 49 is zero. Equations 50 and 53 also yield $P = 0$ because $P^+ = P^-$ and $|E|_{\min} = 0$.

As an example involving the opposite extreme of reflection-coefficient magnitude, consider a lossless line terminated in its characteristic resistance. The voltage and current at each point of the line are in phase and have the magnitudes $|E^+|$ and $|I^+|$. The transmitted power is thus $|E^+|\,|I^+|$ as calculated by either Eq. 49 or 50. Equation 53 yields this same value because $|E|_{\max}$ and $|E|_{\min}$ each equal $|E^+|$ and $|E^+|/R_c$ is equal to $|I^+|$.

In general, lines operate under conditions intermediate to those of complete absorption and complete reflection of the incident power by the load. Where the objective is to transfer power from source to load, as little reflection as possible is ordinarily desired. The existence of a returning wave results, for a given net power transfer, in unnecessarily high values of voltage and current at the maxima of the standing-wave patterns, and these large voltages and currents produce several undesirable effects.

One undesirable effect of high standing-wave ratios is the reduction of the *power capacity* of the line; that is, reduction of the maximum power that can be transmitted by the line. Ordinarily the power that may safely be applied to the line is limited by the voltage that may appear between conductors without danger of insulation breakdown. Let $|E|_{\text{lim}}$ be the limiting or maximum permissible rms value of the voltage between conductors. If the line is operated with a standing-wave ratio of unity, the voltage between conductors is the same in magnitude at all points of the line, and by Eq. 53 the maximum power that can be transmitted is

$$P_{\max} = \frac{|E|_{\text{lim}}^2}{R_c}$$

If the standing-wave ratio is ρ, $|E|_{\min}$ in Eq. 53 is equal to $|E|_{\max}/\rho$, and $|E|_{\max}$ must not exceed $|E|_{\text{lim}}$, so that the maximum power becomes

$$P_{\max} = \frac{|E|_{\text{lim}}^2}{\rho R_c} \tag{54}$$

Thus standing waves cause a reduction of power capacity, the maximum power that can be delivered from source to load being inversely proportional to the standing-wave ratio.

Additional undesirable effects of high standing-wave ratios are discussed in Arts. 13 and 15.

13. Losses and Attenuation. In the preceding articles, the discussions of transmission-line behavior are based upon the assumption that the lines are lossless. This assumption is only approximately true, because a small amount of power is always dissipated along the line. Losses in coaxial lines include *conductor loss*, or power lost as heat because of current flow through the resistance of the conductors, and *dielectric loss*, or power loss because of currents flowing in imperfect insulating material between conductors. Parallel-wire lines are subject to these same losses and (if the electric and magnetic fields surrounding the conductors are not confined by a conducting shield) have additional losses occasioned by radiation and by interaction of the fields with adjacent objects and circuits.

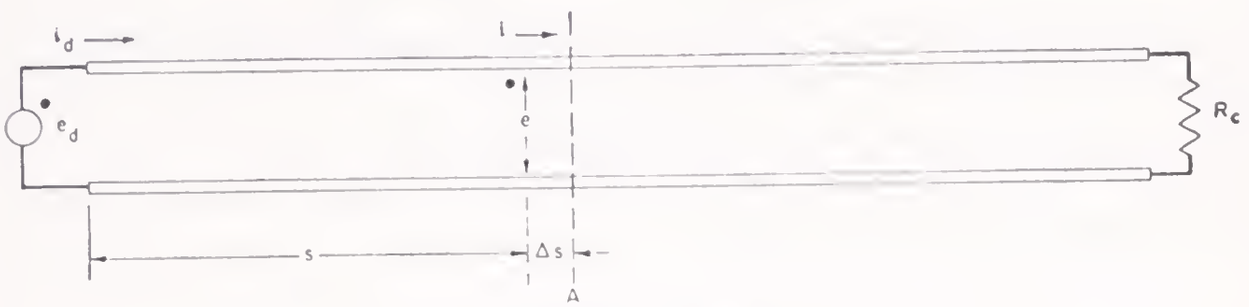
For a radar transmission line not more than a few waves long, the power dissipated is ordinarily a negligible fraction of the total power transmitted along the line, and calculations that ignore losses are essentially correct. For lines of great length, however, the total loss of power from all parts of the line may be appreciable and thus may affect values of voltage, current, and impedance along the line. Where the purpose of the line is to deliver power from a source to a load, even small losses represent an imperfection of the line, and every precaution must be taken to minimize the loss and to deliver to the load the largest possible fraction of the source power.

The major effect of small power losses is illustrated by Fig. 24, in which is shown a long section of line carrying a single traveling wave of voltage and current. In most respects, the wave propagation occurs in the same way as on a lossless line—the voltage and current waves travel in phase at a speed $1/\sqrt{lc}$ and with an amplitude ratio $|E|/|I|$ equal to R_c. However, as the wave traverses the line, its power $P = |E|\,|I|$ is continually decreased by the losses. As shown in the figure, the amplitude of the voltage and current waves also decrease or *attenuate* as the distance from the source increases.

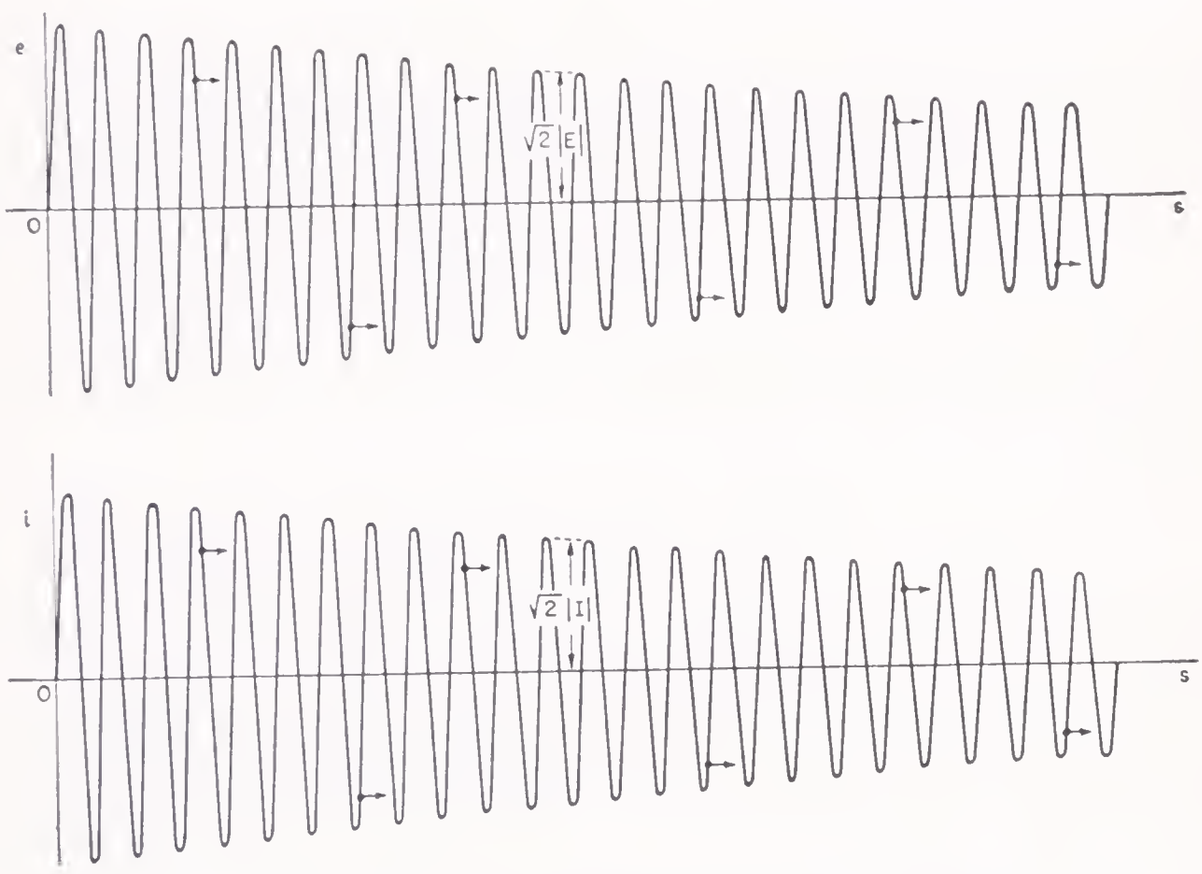
If the attenuation is so great that the decrease of amplitude in one wavelength is appreciable, additional effects of losses may become noticeable. The losses may introduce a phase angle between the waves of voltage and current (corresponding to a complex value for the characteristic impedance) and may make the speed of propagation less than $1/\sqrt{lc}$ and a function of frequency.

In this text only lines with *small loss* are considered, that is, lines for

which the attenuation is appreciable only if observations are made over a distance comprising many wavelengths and for which other effects of loss can be ignored. Radio-frequency lines having greater losses are almost never encountered in radar for the following reason: Because of the high frequency of operation, the wavelength is very short—of the



(A) CIRCUIT DIAGRAM



(B) VOLTAGE AND CURRENT WAVES AT A PARTICULAR INSTANT

FIG. 24. Traveling wave on line with small loss.

order of a few inches. Lines are designed to have reasonably small loss in the entire length employed, which may be several tens of feet. Thus the amount of the loss in a single wavelength is negligible.

For lines of small loss, the basic assumptions are made that the voltage and current of a traveling wave are in phase and that the values of propagation speed and characteristic resistance are not altered by the losses. For such a line, the amount of the attenuation may be calculated simply in terms of the line parameters. These parameters include not

only l and c (the inductance and capacitance of a unit length of line which determine R_c and the speed of propagation) but also the loss parameters r (the conductor resistance for a unit length of line) and g (the shunt conductance of the dielectric in a unit line length).

Consider first the attenuation in a line traversed by a single traveling wave, as in Fig. 24. At some point A, the voltage and current have the rms values $|E|$ and $|I|$, and these values do not vary appreciably over a short section of line of length Δs adjacent to A. The conductor resistance in the Δs section is $r\Delta s$, and the shunt conductance is $g\Delta s$. The power loss of the section is the sum of the power losses in the conductors and dielectric and is very nearly

$$|I|^2 r\Delta s + |E|^2 g\Delta s = |E|\,|I|\,\frac{r}{R_c}\,\Delta s + |E|\,|I| g R_c \Delta s$$
$$= \left(\frac{r}{R_c} + gR_c\right) P\Delta s$$

This power loss represents a decrease in the transmitted power that occurs between the input and output terminals of the Δs section. Thus

$$\Delta P \approx -\left(\frac{r}{R_c} + gR_c\right) P\Delta s \tag{55}$$

where ΔP stands for the change of P in the section of length Δs. Equation 55 becomes more nearly exact as the length Δs decreases, and in the limit

$$\frac{dP}{ds} = -\left(\frac{r}{R_c} + gR_c\right) P \tag{56}$$

where dP/ds is the rate of change of the power with distance, or $-dP/ds$ is the power loss per unit length of line.

The procedure used in the derivation of Eq. 55 is often referred to as a *perturbation method*, because it is based upon introducing losses as a small change or perturbation of the operating conditions of a lossless line for which calculations have previously been made. This general type of calculation is used again in Chaps. VIII and IX in the calculation of losses in waveguides and cavity resonators.

Equation 56 specifies that the rate of decrease of power with distance along the line is proportional to the value of the power. In Art. 4 of Chap. II, it is shown that a curve having a slope proportional to its ordinate is an exponential. Therefore, the power of a traveling wave on a uniform transmission line decreases exponentially with distance along the line. In like manner, the voltage and current magnitudes decrease exponentially and may be expressed as

$$|E| = |E_d|\epsilon^{-\alpha s}, \qquad |I| = |I_d|\epsilon^{-\alpha s} \tag{57}$$

where $|E_d|$ and $|I_d|$ are values of $|E|$ and $|I|$ at the source (where $s = 0$, see Fig. 24A) and α is a constant, called the *attenuation constant* of the line. Because power is proportional to the square of voltage or current, the corresponding exponential relation for power is

$$P = P_d \epsilon^{-2\alpha s} \tag{58}$$

where P_d is the value of P at $s = 0$. The derivative of Eq. 58 is

$$\frac{dP}{ds} = -2\alpha P_d \epsilon^{-2\alpha s} = -2\alpha P \tag{59}$$

and a comparison of this relation with Eq. 56 yields the value of the attenuation constant:

$$\alpha = \frac{1}{2}\left(\frac{r}{R_c} + gR_c\right) \tag{60}$$

The effect of a line of length d upon a traveling wave is, according to Eq. 58, to reduce the power of the wave in a definite ratio. If the input power is P_d and the load power is P_L, the attenuation of the line may be specified by the power ratio P_L/P_d. Alternatively and more conveniently, the attenuation may be specified by the value of αd associated with the line. Because P_L/P_d is equal to $\epsilon^{-2\alpha d}$,

$$\alpha d = \frac{1}{2} \ln \frac{P_d}{P_L}$$

The quantity αd is called the *attenuation in nepers* of a line of length d, and the attenuation constant α is thus measured in nepers per unit of length.

Observe that the neper, like the decibel (see Art. 16, Chap. VI) is a logarithmic measure of power ratio. Line attenuation can be measured in decibels as well as in nepers, and a definite relation exists between the two units. From the definitions in Art. 16, Chap. VI, the attenuation in decibels of a line of length d is

$$\text{Attenuation in decibels} = 10 \log \frac{P_d}{P_L} \tag{61}$$

The natural logarithm of P_d/P_L is 2.303 times the common logarithm of P_d/P_L, and therefore the number expressing an attenuation in nepers is $\frac{1}{2} \times 2.303/10 = 0.1151$ times the number expressing the same attenuation in decibels; that is, the neper is the larger unit, and the ratio of the units is 8.686. The logarithmic units *neper* and *decibel* are convenient for use in attenuation calculations because, in terms of these units, the total attenuation of several line sections or other devices in cascade is the sum of the attenuations of the component devices.

The magnitudes of the voltage and current of a traveling wave on the line of Fig. 24*A* decrease exponentially with distance in accordance with Eq. 57. At the same time, the phase angles change in proportion to distance exactly as for waves on a lossless line. The combined effects of attenuation and propagation speed can be expressed by combining the exponential attenuation factor from Eq. 57 with the exponential phase factor $\epsilon^{-j\beta s}$. Thus

$$E = E_d\epsilon^{-\alpha s}\epsilon^{-j\beta s}, \qquad I = I_d\epsilon^{-\alpha s}\epsilon^{-j\beta s} \tag{62}$$

where E and I are the complex numbers representing line voltage and current and E_d and I_d are values of E and I at the source. Equations 62 may be written

$$E = E_d\epsilon^{-(\alpha+j\beta)s}, \qquad I = I_d\epsilon^{-(\alpha+j\beta)s}$$

or

$$E = E_d\epsilon^{-\gamma s}, \qquad I = I_d\epsilon^{-\gamma s} \tag{63}$$

where

$$\gamma = \alpha + j\beta \tag{64}$$

The quantity γ is called the *propagation constant* of the line, the real part of γ being the attenuation constant and the imaginary part being the phase constant.

Thus far, small losses on a line terminated in its characteristic resistance have been considered, and the wave that travels from source to load has been found to attenuate exponentially as it progresses. Also of interest is the effect of losses upon transmission-line waves when the terminating impedance is different from R_c. Such a line is represented in Fig. 25*A*, the special case of a resistive load impedance $3R_c$ being assumed in the figure.

The total voltage and current for the line of Fig. 25*A* may be divided into two traveling-wave components, exactly as is done for lossless lines. Because of the losses, the forward-going wave is attenuated as it moves toward the load and the returning wave is attenuated as it progresses toward the source. Thus, at some instant, the voltages of the two waves are as indicated in Fig. 25*B*. The returning wave is in phase with the forward-going wave at the load and has an amplitude one-half the forward-wave amplitude because, for $Z_L = 3R_c$, the reflection coefficient is $0.5 + j0$.

The sum of the two traveling waves yields the standing-wave pattern of Fig. 25*C*. Note that the standing-wave ratio of the pattern is 3 near the load (corresponding to $|\Gamma_L| = 0.5$) but it is less than 3 at points far removed from the load because of the increased amplitude of the forward-going wave and the decreased amplitude of the returning wave. Attenuation in a uniform line always causes the standing-wave ratio to decrease

toward unity as the point of observation is moved from the location of a mismatch toward the source. By the use of considerable lengths of high-attenuation line, the impedance that the source faces can be made

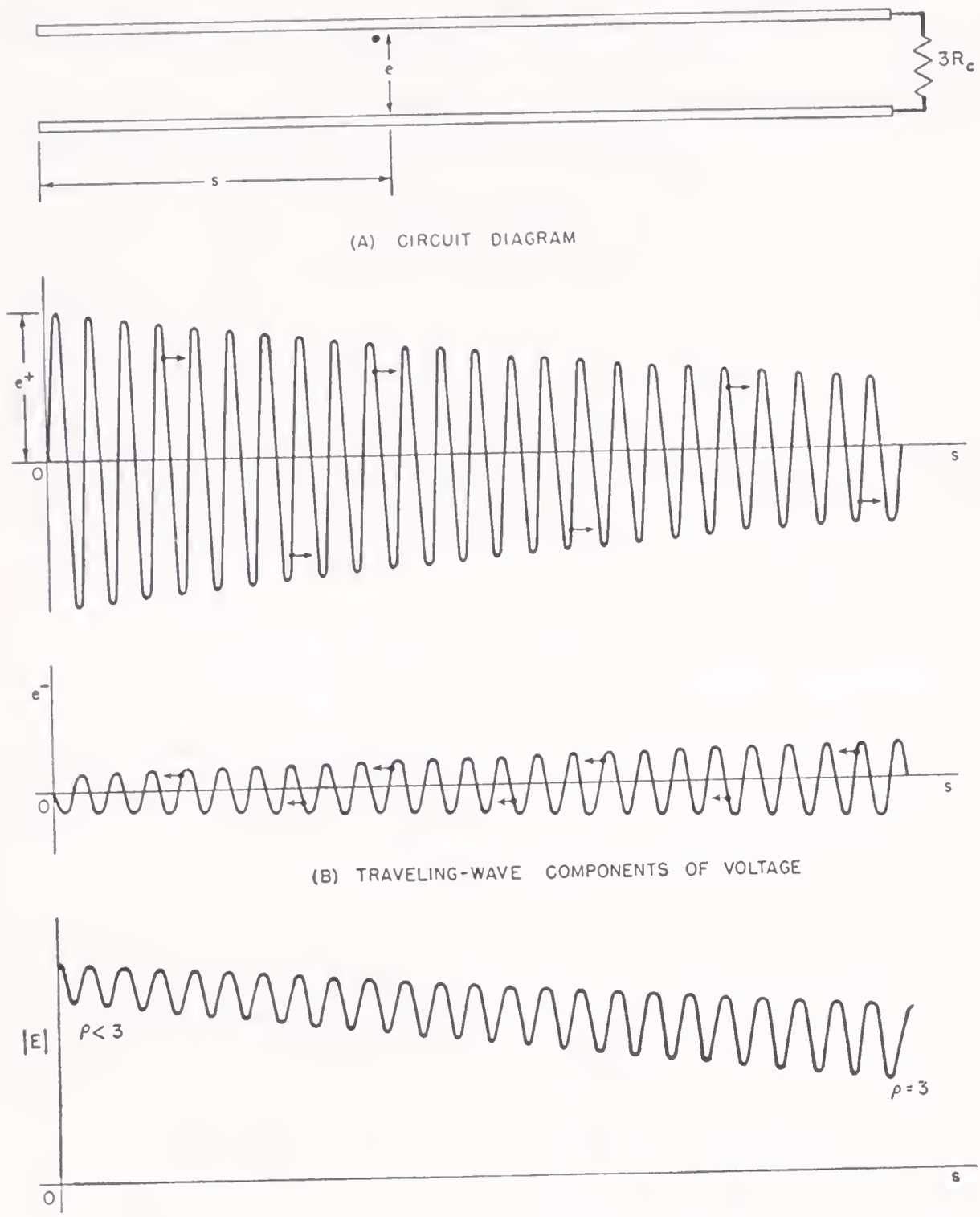


FIG. 25. Effect of losses upon standing-wave pattern.

equal to the characteristic impedance regardless of the value of the load impedance.

Of special interest is the effect of standing waves upon the efficiency with which power is transferred by a line from a source to a load. For a line of length d and attenuation constant α operating with a matched

load, the efficiency η_0 is

$$\eta_0 = \frac{P_L}{P_d} = \epsilon^{-2\alpha d} \tag{65}$$

where P_d is the input power and P_L the output power, and the relation of these quantities is specified by Eq. 58 with $s = d$.

If the line is terminated in an impedance different from R_c, Eq. 50, p. 511, can be used to express P_L and P_d in terms of the power values associated with the forward-going and returning waves. (As noted in Art. 12, Eq. 50 is valid only for a line having a resistive characteristic impedance, but this condition is met by the small-loss lines considered here.) The efficiency η of a line with any value of terminating impedance may thus be expressed as

$$\eta = \frac{P_L^+ - P_L^-}{P_d^+ - P_d^-} \tag{66}$$

The power quantities P_L^+ and P_d^+ associated with the forward-going wave are related by

$$P_L^+ = P_d^+ \epsilon^{-2\alpha d} = \eta_0 P_d^+$$

Similarly, for the power of the returning wave,

$$P_d^- = P_L^- \epsilon^{-2\alpha d} = \eta_0 P_L^-$$

Furthermore, the ratio P_L^-/P_L^+ is equal to the square of the ratio of the corresponding voltage magnitudes, and therefore

$$P_L^- = |\Gamma_L|^2 P_L^+$$

These relationships permit all the quantities in the fraction of Eq. 66 to be expressed in terms of P_d^+. The line efficiency is therefore

$$\eta = \frac{\eta_0 P_d^+ - \eta_0 |\Gamma_L|^2 P_d^+}{P_d^+ - \eta_0^2 |\Gamma_L|^2 P_d^+} = \eta_0 \frac{1 - |\Gamma_L|^2}{1 - \eta_0^2 |\Gamma_L|^2} \tag{67}$$

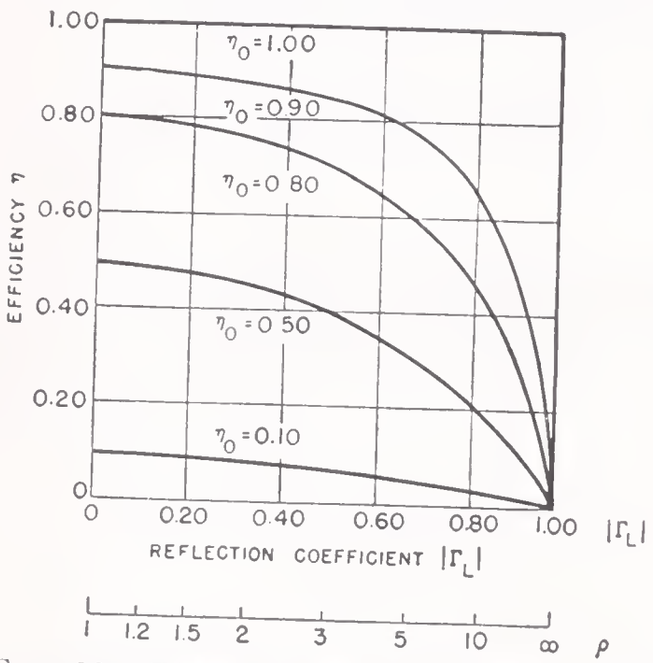


FIG. 26. Effect of standing waves upon transmission-line efficiency. (The characteristic impedance is assumed to be a resistance.)

By use of Eq. 67, η is plotted in Fig. 26 as a function of $|\Gamma_L|$ for several values of η_0. A nonlinear abscissa scale of ρ is also provided. From these curves, it is evident that efficiency is reduced by the presence of standing waves. The loss of efficiency is not large, however, unless the standing-wave ratio is greater than about 3.

14. Impedance of Lines Having Small Losses. The operating characteristics of many transmission-line devices depend upon impedance relations in line sections less than one wave long. In such short line sections, the effect of attenuation is usually negligible, and impedance calculations can be made by the methods of Arts. 8 to 10 as though the lines were entirely lossless. Examples of such calculations are given in Art. 15.

Sometimes, however, impedance relations are of interest in a line many waves long. The effects of attenuation are then important even though the loss in a single wavelength is negligible, and the change of standing-wave ratio with distance from a point of mismatch must then be considered. For a lossless line the reflection coefficient at a distance s from the load differs from the load reflection coefficient Γ_L because of a change of angle and is equal to $\Gamma_L \epsilon^{-j2\beta s}$ (see Eq. 38, p. 496). For a line with losses, not only the angle, but also the magnitude of the reflection coefficient is a function of position. If Eq. 63, p. 518, is written in terms of a distance s measured from the load instead of from the source, the reflection coefficient of an attenuating line may be computed as follows:

$$E^+ = E_L^+ \epsilon^{\gamma s}$$
$$E^- = E_L^- \epsilon^{-\gamma s}$$

and therefore

$$\Gamma = \frac{E^-}{E^+} = \frac{E_L^-}{E_L^+} \epsilon^{-2\gamma s}$$

or

$$\Gamma = \Gamma_L \epsilon^{-2\gamma s} = \Gamma_L \epsilon^{-2\alpha s} \epsilon^{-j2\beta s} \tag{68}$$

The factor $\epsilon^{-2\alpha s}$ in Eq. 68 indicates the effect of attenuation on the magnitude of Γ, and the factor $\epsilon^{-j2\beta s}$ indicates the effect of propagation speed upon the phase of Γ.

The three-step process specified by Eqs. 37 to 39 in Art. 8 for impedance calculations in lossless lines can be applied also to lines with small loss, provided Eq. 68 rather than Eq. 38 is used to relate reflection coefficients at different points along a line. Equations 37 and 39 relating impedance to reflection coefficient are unchanged. Therefore the Smith chart and z-plane diagram, which are graphical representations of the impedance-to-reflection-coefficient relation, may be used in the calculations. In fact, all the procedures for using the Smith chart presented for lossless lines may be applied to lines with small loss, with only one change: To determine the reflection coefficient at one point of a line from that at another, the representative point is moved, not along a circle of constant ρ, but rather along a spiral about the center of the Smith chart. The radius of the spiral decreases for a clockwise rotation of the representative point corresponding to a motion toward the generator of the point of observa-

tion on the line. The rate at which the radius of the spiral changes is specified by the factor $\epsilon^{-2\alpha s}$ in Eq. 68.

As an example of impedance relations in a line with small loss, consider the line of Fig. 27*A* for which the characteristic resistance is 50 ohms, the speed of propagation is 3×10^8 m per sec, the attenuation is 0.5 db per m, and the length is 2.025 m. The input impedance is to be determined for a load impedance of $100 + j0$ and an operating frequency of 3,000 Mcps. The wavelength, determined from the speed and frequency of the waves, is 10 cm; the line is therefore $20\frac{1}{4}$ waves in length.

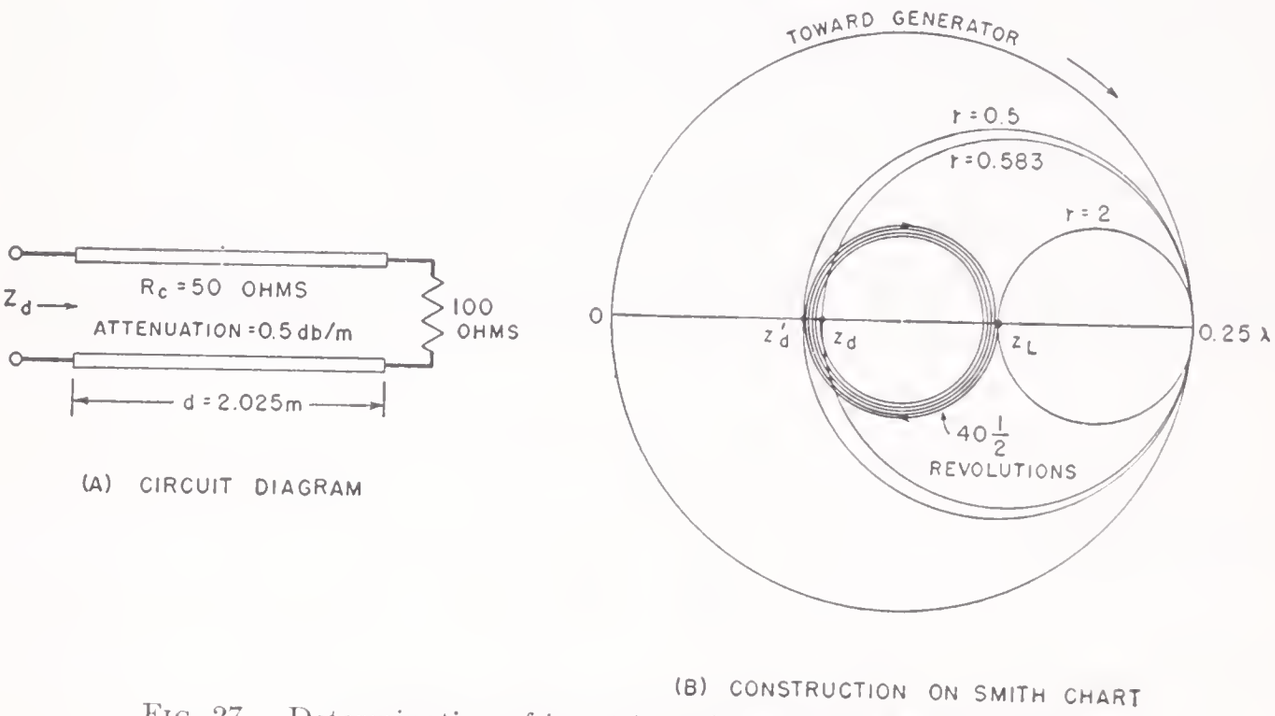


FIG. 27. Determination of input impedance of line with small loss.

The load impedance is represented by the point z_L in Fig. 27*B* at a per-unit impedance value of $2 + j0$. The input impedance z_d is obtained by a clockwise rotation through an angle corresponding to $20\frac{1}{4}$ wavelengths ($40\frac{1}{2}$ revolutions in the chart). If the line were lossless, the result would be the same as that obtained by a rotation of one-quarter wavelength, and the per-unit input impedance (represented by z_d' in Fig. 27*B*) would be $0.5 + j0$. Because of the losses, however, the path of the representative point is, as shown in the figure, a closely wound spiral ending in point z_d at the per-unit value of the actual input impedance.

Point z_d is located on the same radial line from the center of the chart as z_d', but the distance from the center to z_d is, by Eq. 68, $\epsilon^{-2\alpha d}$ times the distance from the center to z_d'. The factor $\epsilon^{-2\alpha d}$ can be evaluated by first converting the attenuation in decibels to attenuation in nepers (see Art. 13), as follows:

$$\epsilon^{-2\alpha d} = \epsilon^{-2\times(0.5/8.686)\times 2.025} = \epsilon^{-0.233} = 0.793$$

Alternatively, the ratio can be determined directly from the definition of the decibel as

$$10^{-2\times(0.5/20)2.025} = 10^{-0.10125} = 0.793$$

From the point z_d thus located, the input impedance is read as $0.583 + j0$ per unit, or $29.15 + j0$ ohms. Note that the effect of losses is, as expected, to produce a value of input impedance more nearly equal to the characteristic resistance than would be obtained with a lossless line.

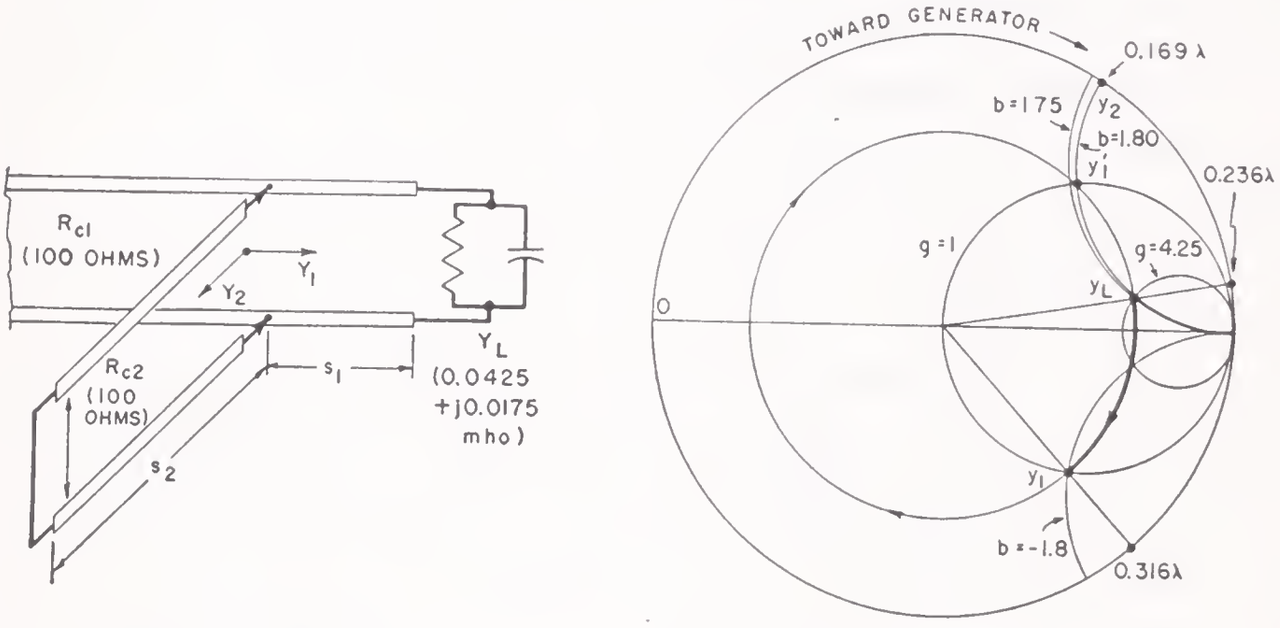
15. Impedance-changing Devices. The performance of a system employing a transmission line to deliver power from a source to a load depends greatly upon impedance relations at the ends of the line. At the source end, the input impedance of the line affects the power output (and often the frequency stability) of the source, and at the load end the terminating impedance determines the standing-wave ratio at which the line operates. Usually two devices for changing impedance, which may be called *impedance transformers*, are employed, one at either end of the line, to secure desired impedance relationships. An impedance transformer near the load permits adjustment for nearly unity standing-wave ratio over the major part of the line, regardless of the impedance of the load itself. A second transformer near the source makes possible operating the source into any desired load impedance when the input impedance of the line is the value R_c associated with a flat line (a line for which $\rho = 1$).

Although transmission-line systems occasionally employ only one impedance transformer, which is adjusted for proper loading of the source without regard to standing waves on the line, use of two transformers and flat lines is generally preferable. As explained in Art. 12, elimination of standing waves makes the *power capacity* of the line a maximum, and as explained in Art. 13, operation at a low standing-wave ratio also provides maximum *efficiency* of power transmission. A third reason for maintaining low standing-wave ratio is the *frequency sensitivity* of the line. If the line is many wavelengths long, marked variations of the input impedance result from even small shifts of frequency unless the standing-wave ratio is very near to unity. Thus low standing-wave ratios are required in broad-band systems. Often also, impedances nearly independent of frequency and, therefore, low-standing-wave ratios are necessary for satisfactory operation of vacuum-tube power sources (see Art. 10 of Chap. X and Art. 10 of Chap. XI).

Single-stub Transformer. A form of impedance transformer often employed with parallel-wire lines is the single-stub transformer, a device that consists of a short section of transmission line connected in parallel with the main line and terminated in a short circuit. In order that any impedance-changing device may provide a complete range of control of

impedance, it must be adjustable in two ways so that the resistance and reactance components of impedance may be adjusted independently. For the single-stub transformer, the two adjustments are the length of the stub and the position of the stub along the main line.

Consider the use of a stub located near the load end of a line, as in Fig. 28*A*, for the purpose of eliminating standing waves on the main line



(A) CIRCUIT DIAGRAM (B) CONSTRUCTION ON SMITH CHART

FIG. 28. Example illustrating use of single-stub transformer.

between the stub and the generator. Let Y_1 be the input admittance of the line section to the right of the stub, and let Y_2 be the input admittance of the stub. Because Y_2 has no conductance component, the procedure for making the sum of Y_1 and Y_2 equal to the characteristic conductance of the main line is as follows: (1) Locate the stub at a point where the conductance component of Y_1 is equal to $G_c = 1/R_c$. (2) Adjust the length of the stub to make the total susceptance of the sum of Y_1 and Y_2 equal to zero.

As an example of this procedure, let it be required to determine the location and length of the stub required to obtain a flat line to the left of the stub when the load admittance has the value indicated in Fig. 28*A*. Calculations may be carried out conveniently with the aid of the Smith chart as indicated in Fig. 28*B*. The point representing the per-unit load admittance y_L is located in Fig. 28*B* at

$$y_L = \frac{Y_L}{G_{c1}} = Y_L R_{c1} = (0.0425 + j0.0175)100 = 4.25 + j1.75$$

and corresponds to the point at 0.236λ on the wavelength scale. A clockwise rotation on a constant-ρ circle (losses are negligible in the short line lengths considered) to the $g = 1$ circle yields the value $y_1 = 1 - j1.80$ as the per-unit admittance of the line at the stub location. Since y_1 cor-

responds to 0.316λ on the wavelength scale, the stub is located a distance $s_1 = (0.316 - 0.236)\lambda = 0.08\lambda$ from the load. The value of stub admittance y_2 that makes the total susceptance zero at the stub is $y_2 = 0 + j1.80$, and the corresponding value on the wavelength scale is 0.169. A shorted-end stub with the required input admittance has a length

$$s_2 = (0.169 + 0.25)\lambda = 0.419\lambda$$

An alternative location for the stub is at the point corresponding to $y_1' = 1 + j1.80$. Use of this point places the stub at a greater distance from the load than does the point first chosen and therefore may be less desirable.

Double-stub Transformer. The single-stub transformer is not suitable for use with coaxial lines because of the difficulty of building a coaxial stub that can be adjusted in position. An alternate impedance-changing device, the double-stub transformer, provides two independent adjustments through use of two stubs, each stub being fixed in position and adjustable in length. Constructional details for a coaxial-line device of this form are indicated in Fig. 29. The length of each stub is adjusted by means of a movable plunger that makes contact with both the inner and the outer conductor and thus provides a short-circuit termination. The stubs may be separated by any distance less than a half wavelength; usually the separation distance is three-eighths wavelength.

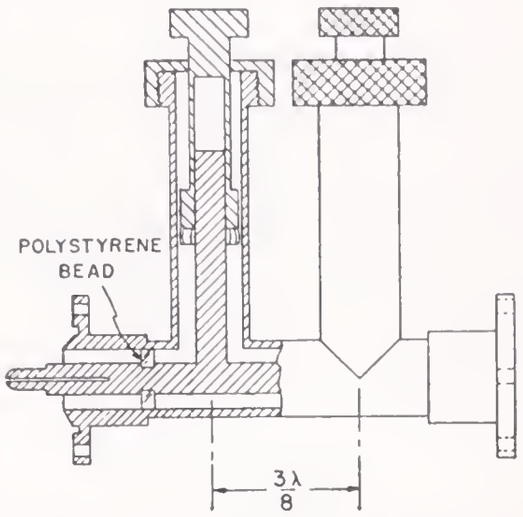
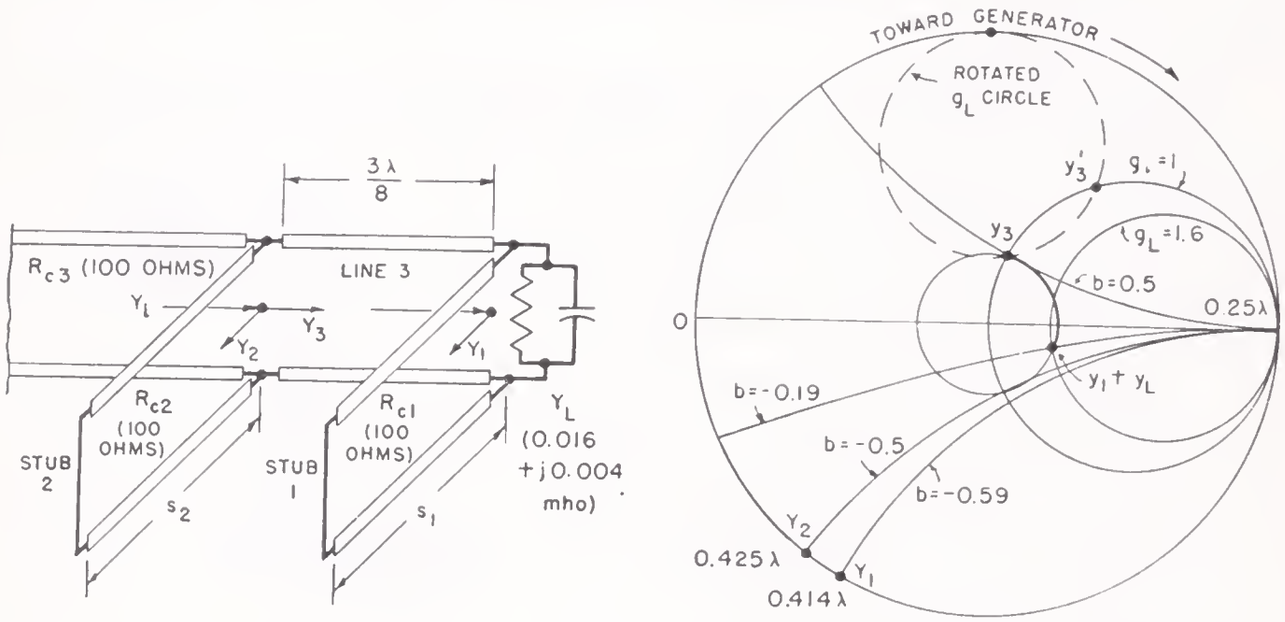


FIG. 29. Double-stub transformer.

Admittance relations in the double-stub transformer may be explained by reference to Fig. 30*A*, which for convenience in drawing depicts a parallel-wire system. Admittance Y_L at the output terminals of the transformer may be the admittance of a physical load device or the input admittance of a line section that connects the transformer to the load. Admittances Y_1 and Y_2 are the admittances of stubs 1 and 2, respectively; Y_3 is the input admittance of line 3 (the line section that connects the two stubs); and Y_i is the value of load admittance for the left-hand line section to be secured by adjustment of the transformer.

Calculation of the stub lengths for a desired value of Y_i depends upon the fact that each stub can alter only the susceptance component of the admittance at its terminals. Thus the desired susceptance component of Y_i can be secured by means of stub 2, but the desired conductance component must be obtained through adjustment of stub 1. The susceptance of stub 1 is adjusted so that the total load admittance $Y_1 + Y_L$ for line 3

provides the desired conductance component of admittance at the input terminals of line 3. In terms of the Smith chart, stub 1 must be adjusted so that Y_3 is represented by some point on the circle of conductance g_i, where g_i is the desired conductance component of Y_i. Similarly, $Y_1 + Y_L$, the load admittance for line 3, must have the same per-unit conductance g_L as does Y_L and must be represented by a point on the g_L circle. The object, then, is to adjust stub 1 so that $Y_1 + Y_L$ is represented by a particular point on the g_L circle which, when rotated through a clockwise angle corresponding to the length of line 3, lies at some point



(A) CIRCUIT DIAGRAM (B) CONSTRUCTION ON SMITH CHART

FIG. 30. Example illustrating use of double-stub transformer.

on the g_i circle. Stub 2 is then adjusted to satisfy the susceptance requirement of Y_i.

As an example of the use of this calculation procedure, assume that the stub separation distance and the load admittance have the values indicated in Fig. 30*A* and that the stub lengths are to be chosen to yield a flat line to the left of stub 2. The per-unit value of the load admittance is $y_L = 1.6 + j0.4$, and thus the point representing the per-unit admittance $y_1 + y_L$ must lie somewhere on the $g_L = 1.6$ circle in Fig. 30*B*. The *possible* values of per-unit input admittance y_3 for line 3 are obtained by a clockwise rotation of each point of the g_L circle through an angle corresponding to the $\frac{3}{8}$ wavelength of line 3. By this rotation, the dashed circle in Fig. 30*B* is obtained. The desired value of admittance y_3 must lie on the circle $g_i = 1$ indicated in Fig. 30*B*, g_i being the conductance component of the per-unit admittance $y_i = 1 + j0$ required to eliminate standing waves on the left-hand line section. Observe that the circles indicating desired and possible values for y_3 intersect at the two points labeled y_3 and y_3' in the figure. Thus there are two possible ways to

adjust the stubs to eliminate standing waves, one corresponding to each of the points of intersection.

For the point y_3, the per-unit admittance value is $1.0 + j0.5$, and thus in order that $y_1 = y_2 + y_3$ may be $1 + j0$, the admittance y_2 of stub 2 must be $-j0.5$, corresponding to a length of

$$0.425\lambda - 0.25\lambda = 0.175\lambda$$

The corresponding value of the admittance $y_1 + y_L$ is obtained from the chart by rotating the y_3 point $\frac{3}{8}\lambda$ counterclockwise along a constant-ρ circle to the $g_L = 1.6$ circle. The point $y_1 + y_L$ thus located is at $1.6 - j0.19$, and thus the admittance of stub 1 is

$$y_1 = (y_1 + y_L) - y_L = (1.6 - j0.19) - (1.6 + j0.4) = -j0.59$$

and the corresponding stub length is

$$0.414\lambda - 0.25\lambda = 0.164\lambda$$

If g_L, the per-unit conductance of the load in Fig. 30*A*, is increased to 2, the rotated g_L circle in Fig. 30*B* becomes tangent to the g_i circle, and only one adjustment of the stubs is possible. For values of g_L greater than 2, the rotated g_L circle does not intersect the g_i circle, and a matched termination cannot be obtained for any adjustment of the stub lengths. Standing waves can be eliminated only by moving the transformer along the line to a point where g_L is less than 2.

The quantitative statement of the limitations of the double-stub transformer with arbitrary stub separation is that the product of the per-unit load conductance and the desired per-unit input conductance cannot exceed the square of the cosecant of the angular distance between stubs. That is, if d is the distance between stubs,

$$g_i g_L \leqq \csc^2 \beta d \tag{69}$$

The proof of Eq. 69 may be made by reference to Fig. 31.

Figure 31 represents a circle of constant conductance drawn for an arbitrary value of g_i with center at N and radius $n = 1/(1 + g_i)$. A similar circle for g_L rotated *clockwise* through the angle $2\beta d$ to correspond to a stub separation βd is shown with center M and radius $m = 1/(1 + g_L)$. Because the two circles are tangent at point P, the diagram represents conditions when g_L has the largest value for which the specified conductance g_i can be produced by a transformer using the separation angle βd. The radius of the outside circle of the Smith chart is unity. Therefore, the sides of the triangle OMN have the lengths indicated in the figure, and by the law of cosines

$$(m + n)^2 = (1 - m)^2 + (1 - n)^2 - 2(1 - m)(1 - n) \cos 2\beta d$$

By expanding the quantities in parentheses, combining terms and rearranging the result, this equation may be changed to

$$\frac{1}{mn} - \frac{1}{m} - \frac{1}{n} = \frac{1 + \cos 2\beta d}{1 - \cos 2\beta d} \tag{70}$$

Expressing m in terms of g_L and n in terms of g_i and making use of the trigonometric identity

$$\tan^2 \frac{\theta}{2} = \frac{1 - \cos \theta}{1 + \cos \theta}$$

in Eq. 70 yields

$$(1 + g_L)(1 + g_i) - (1 + g_L) - (1 + g_i) = \cot^2 \beta d$$

Thus

$$g_i g_L = \cot^2 \beta d + 1 = \csc^2 \beta d$$

and because Fig. 31 is drawn for the maximum value of g_L, Eq. 69 is proved.

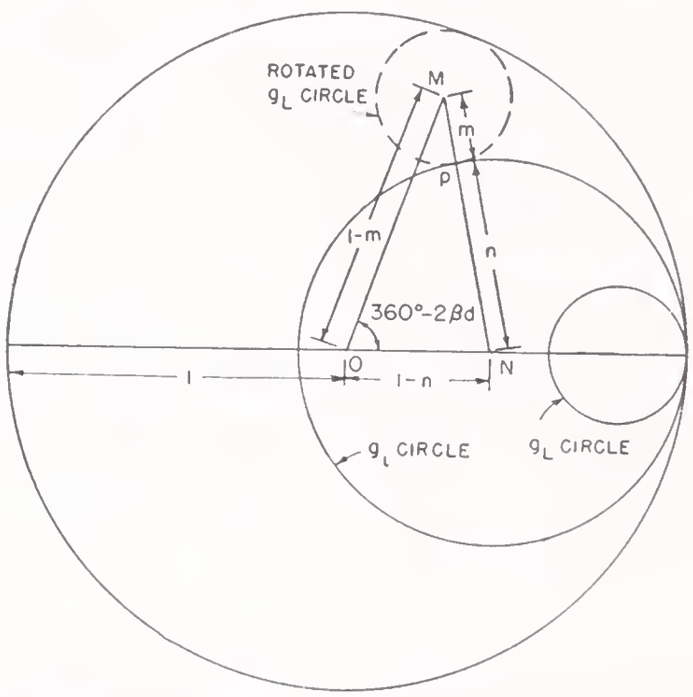


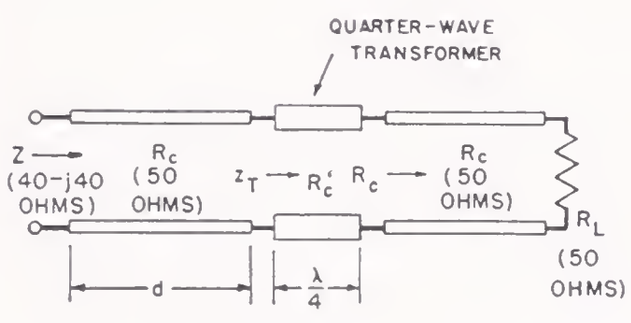
FIG. 31. Diagram used in proof of Eq. 69.

Equation 69 indicates that the range of conductance values with which the double-stub transformer can operate is least if the stub separation is a quarter wave and increases if the separation distance is decreased toward zero or increased toward a half wave. Ordinarily separation distances less than a quarter wave are not possible because of the space requirements of the stubs. Thus spacings between a quarter wave and a half wave are necessary, and a distance of three-eighths wavelength is ordinarily used.

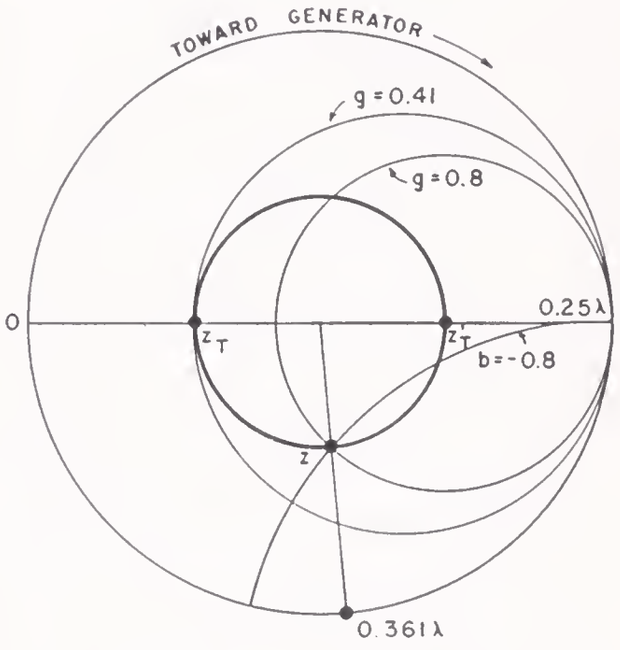
Quarter-wave Transformer. The single-stub and double-stub transformer, as presented above, are adjustable devices that may be reset

whenever the impedance of the load or the impedance required by the source changes. Often the terminal impedances are essentially fixed, and no adjustment of the transformers is required subsequent to the original choice of dimensions in manufacture. A particularly simple form of non-adjustable impedance-changing device, suitable for use with either parallel-wire or coaxial lines, is the *quarter-wave transformer*. This device comprises a section of line a quarter wave in length of different characteristic resistance from the rest of the line. By proper choice of the location and characteristic resistance of the transformer, arbitrarily assigned values of input impedance can be obtained. Usually the characteristic resistance of the transformer is less than that of the line itself, and the transformer is made by placing a close-fitting metal sleeve over the inner conductor of a coaxial line or over each conductor of a parallel-wire line.

The input impedance of a quarter-wave transformer is shown in Art. 8 (see Eq. 42, p. 499) to be R_c^2/Z_L, where R_c is the characteristic resistance of the transformer and Z_L is the load impedance for the transformer. Thus use of the transformer requires that the device be located so that the angle of the impedance desired at the transformer input terminals is the negative of the angle of the impedance presented to the output terminals of the transformer. The characteristic resistance of the transformer is then adjusted to satisfy the magnitude relationship.



(A) CIRCUIT DIAGRAM



(B) CONSTRUCTION ON SMITH CHART

FIG. 32. Example illustrating use of quarter-wave transformer.

As an example of the use of a quarter-wave transformer, consider the circuit of Fig. 32*A*. A 50-ohm line terminated in its characteristic resistance has a quarter-wave transformer near its input terminals. The characteristic resistance R_c' of the transformer and the distance d between transformer and the input terminals are to be chosen so that the input

impedance is $40 - j40$ ohms, or $0.8 - j0.8$ per unit. Because the load impedance for the transformer is the resistance R_c, the input impedance Z_T of the transformer must also be resistive and must be represented on the Smith chart (see Fig. 32*B*) by a point on the horizontal axis. In order that the desired transmission-line input impedance may be provided, the point representing Z_T must lie on the same ρ circle as the point $z = 0.8 - j0.8$. Thus the circle through z is drawn as in Fig. 32*B*, and the two points of intersection with the horizontal axis z_T and z_T' are located.

Either of these points of intersection may be used to determine R_c' and d, but only point z_T yields a value of R_c' less than R_c and thus permits constructing the transformer by placing a sleeve over a line conductor. The clockwise rotation required to go from z_T to z determines the distance d as 0.361λ, and the impedance at z_T, $0.41 + j0$ per unit, may be used to determine the value of R_c'. Because of the relation $Z = R_c^2/Z_L$ for a quarter-wave transformer, R_c' must be the geometric mean of the impedances 1.0 and 0.41 per unit. Thus R_c' has a per-unit value, referred to the 50-ohm line of $\sqrt{1 \times 0.41} = 0.64$, or R_c' is equal to 32 ohms.

Bandwidth of Impedance-changing Devices. In the above discussions of single-stub, double-stub, and quarter-wave transformers, operation at only a single frequency is considered. The desired impedance relations are not, in general, maintained at other frequencies because of the change in electrical length of line sections with change of wavelength. However, over a bandwidth of a few per cent, the change in wavelength is small, and the resulting errors in impedance values are for many purposes negligible. Bandwidth of impedance transformers is of interest in radar because of the extent in frequency of the spectrum of the radio-frequency pulses employed (see Art. 4, Chap. VI) and because of the variability of the frequency of practical transmitters. For many radar applications, the simple transformers discussed above provide adequate bandwidth, provided high standing-wave ratios on electrically long lines are avoided.

Impedance relationships may be controlled over wider bandwidths by the use of specially designed devices. One way of obtaining broad-band operation is by the use of several transformer sections that introduce a change of impedance gradually. For example, several quarter-wave sections of slightly different characteristic resistances may be employed.[1] In general, the greater the number of the cascaded sections and the smaller the discontinuity at each junction, the less sensitive to frequency is the over-all device. A tapered section of line one wave or more in length may be used to join lines of different values of characteristic resistance with little reflection at the junction over a wide frequency band.

[1] J. C. SLATER, *Microwave Transmission*, 56–78 (McGraw-Hill Book Company, Inc., New York, 1942).

16. Transmission-line Capacitance. Methods have been presented in the preceding articles for determining the behavior of a transmission line if the characteristic resistance of the line and the speed of propagation and attenuation of waves on the line are known. Methods have been given also for determining the characteristic resistance, propagation speed, and attenuation in terms of the line parameters l, c, r, and g, the inductance, capacitance, conductor resistance, and dielectric conductance per unit length of the line. There remain to be presented in this chapter the methods of calculating the line parameters from the dimensions and electrical properties of the conductors and dielectric of the line.

In the determination of c, the capacitance per-unit length of line, the conductors of a unit length of line are considered as plates of a capacitor, and the ratio of a charge placed on the conductors to the voltage produced between conductors is calculated. This ratio, which is the capacitance sought, depends upon the conductor dimensions and the nature of the dielectric material between the conductors. Calculation of the ratio requires a knowledge of the *electric field* in the dielectric.

Relationships involving electric-field or magnetic-field quantities are considerably simplified if the rationalized mks system of units is employed.[1] In this system the *m*eter, *k*ilogram, and *s*econd are the basic units of length, mass, and time, and a fourth fundamental unit is added for the specification of electromagnetic quantities. This fourth unit is the coulomb, the unit of electric charge. The ordinary electric-circuit units (watt, ampere, volt, ohm, farad, and henry) are derived from these basic units, and field-quantity units are obtained by combinations of the electric-circuit units with the units of length and time.

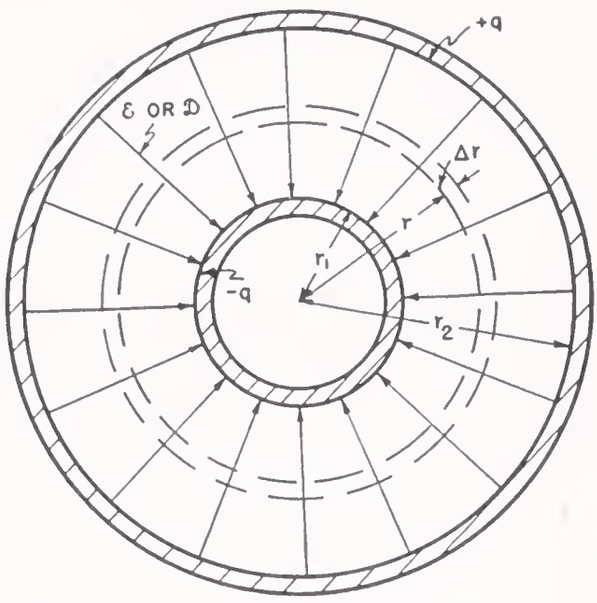


FIG. 33. Cross section of coaxial line showing electric field.

Consider a coaxial-line section of unit length having the cross section indicated in Fig. 33. The radius of the outer surface of the inner conductor is r_1, and that of the inner surface of the outer conductor is r_2. Electric-field relations in this line section are to be determined as one step in the calculation of the line capacitance.

Assume that a positive charge of q coulombs is placed on the outer conductor and an equal-magnitude negative charge is placed on the inner conductor. These charges distribute themselves uniformly over the con-

[1] A. E. KENNELLY, "I.E.C. Adopts the MKS System of Units," *Elec. Eng.*, **54** (December, 1935), 1373–1384.

ductor surfaces and give rise to an electric field in the dielectric between conductors. The field can be described, as shown in this figure, by lines of *electric flux* directed from the positive charges to the negative charges. In the rationalized mks system, one line is drawn from each unit of charge. Thus the total flux Ψ is measured in coulombs, the same unit as the charge. In Fig. 33, Ψ is equal to q coulombs.

Because of the symmetry of the coaxial conductors, the lines of electric flux are directed radially inward and are uniformly spaced both around and along the transmission line. These flux lines form a graphical representation of the *electric-flux density*, a vector quantity directed along the electric-flux lines and having a magnitude equal to the density of the lines. In Fig. 33, the flux-density vector varies in magnitude and direction from point to point of the dielectric. The magnitude is greatest where the flux lines are most concentrated near the inner conductor, and the direction changes with position around the line in such a way that the vector is always directed toward the axis of the cylindrical conductors.

The magnitude $\mathfrak{D}$ of the flux-density vector is measured (in mks units) in coulombs per square meter. For a point on a cylindrical surface of radius r between the two conductors (see Fig. 33), the value of $\mathfrak{D}$ may be calculated as follows: The total flux Ψ penetrating the cylindrical surface of unit length is q coulombs, and the area of the surface is $2\pi r \times 1$. The flux is uniformly distributed over the surface of the cylinder, and thus the flux density is

$$\mathfrak{D} = \frac{\Psi}{\text{area}} = \frac{q}{2\pi r} \tag{71}$$

An alternate way of describing the electric field in Fig. 33 is in terms of the equipotential surfaces—surfaces on which the electric potential is constant. Two equipotential surfaces are the cylindrical surfaces of the conductors that bound the dielectric material. Because of the symmetry of the system, other equipotential surfaces are intermediate cylinders coaxial with the conductors, such as the cylinder of radius r represented by a dashed circle in the figure. The *electric-field intensity* is a vector quantity related to the equipotential surfaces. The direction of the field-intensity vector is perpendicular to the surfaces, and $\mathcal{E}$, the magnitude of the vector, specifies the potential gradient or rate of change of potential with distance along a line perpendicular to the surfaces. In mks units, $\mathcal{E}$ is measured in volts per meter.

If the dielectric material for the line of Fig. 33 is space or any substance for which the dielectric properties are the same in all directions, the direction of the field-intensity vector is everywhere the same as the direction of the flux-density vector, and the lines of electric flux represent the pattern of the field-intensity vector as well as that of the flux-density vector. In

fact, for dielectric material ordinarily employed in transmission lines, the two vectors differ only by a constant factor, and at each point of the dielectric

$$\mathfrak{D} = \varepsilon \mathcal{E} \tag{72}$$

where ε is a constant of proportionality called the *permittivity* of the dielectric. The units of ε are those of $\mathfrak{D}/\mathcal{E}$, or in the mks system,

$$\frac{\text{Coulombs/square meter}}{\text{Volts/meter}} = \frac{\text{coulombs/volt}}{\text{meter}} = \text{farads per meter}$$

The permittivity of space is designated by ε_0 and has the value

$$\varepsilon_0 = 8.854\ \mu\mu f \text{ per m} \tag{73}$$

From Eqs. 71 and 72, the electric-field intensity at any point of the dielectric for the line in Fig. 33 is found to have the magnitude

$$\mathcal{E} = \frac{\mathfrak{D}}{\varepsilon} = \frac{q}{2\pi r \varepsilon} \tag{74}$$

Were this value independent of r, the total voltage between conductors would be simply $\mathcal{E}$ multiplied by the distance between conductors. However, Eq. 74 shows $\mathcal{E}$ to vary inversely with the radial distance r, and thus an integration is required in the calculation of total voltage. Consider the two equipotential surfaces of radii r and $r + \Delta r$ indicated by dashed circles in Fig. 33. If Δr is very small, $\mathcal{E}$ has very nearly a constant value everywhere between these cylinders, and the voltage between the surfaces is approximately $\mathcal{E}\Delta r$, where $\mathcal{E}$ is given by Eq. 74. The total voltage between conductors may thus be found by dividing the total distance into many small increments, such as Δr, and by adding the corresponding small increments of voltage. The limiting value of the sum obtained as the lengths of the incremental distances are decreased is the integral

$$e = \int_{r_1}^{r_2} \mathcal{E} dr = \int_{r_1}^{r_2} \frac{q}{2\pi r \varepsilon} dr = \frac{q}{2\pi\varepsilon} \int_{r_1}^{r_2} \frac{dr}{r}$$

or

$$e = \frac{q}{2\pi\varepsilon} \ln \frac{r_2}{r_1} \tag{75}$$

where e is the voltage between conductors.

The capacitance c of the unit-length line is the ratio q/e, and thus

$$c = \frac{q}{e} = \frac{q}{(q/2\pi\varepsilon) \ln (r_2/r_1)}$$

or

$$c = \frac{2\pi\varepsilon}{\ln (r_2/r_1)} \tag{76}$$

If the dielectric material is air or any other substance having a *dielectric constant* (ratio of ε for the substance to ε_0) of unity, ε may be replaced by $\varepsilon_0 = 8.854 \times 10^{-12}$ farad per m

$$c = \frac{55.6}{\ln (r_2/r_1)} \qquad \mu\mu\text{f per m} \tag{77}$$

The capacitance of a unit length of parallel-wire line can be calculated by similar methods. The electric field is more complicated for the parallel-wire line than for the coaxial line and spreads out in all directions instead of being confined between the two conductors. The nature of

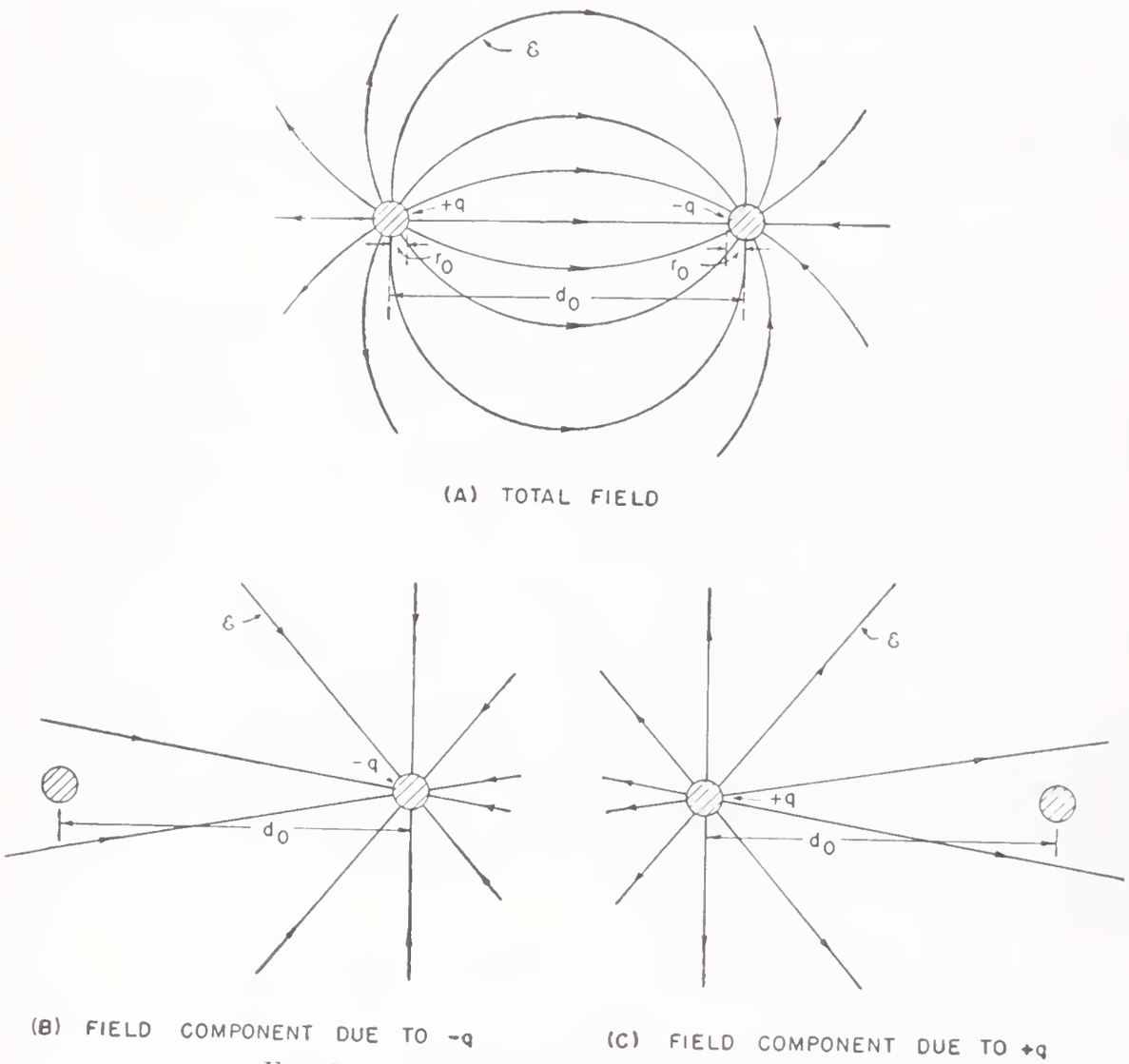


FIG. 34. Electric field of parallel-wire line.

the field is indicated by the sketch in Fig. 34*A* of flux lines in a transverse plane of the parallel-wire line. A simple relation for the capacitance of the line results when d_0, the distance between center lines of the conductors, is large compared with r_0, the radius of each conductor. Charges of $+q$ and $-q$ placed on the conductors of a unit-length line then have little effect upon each other and distribute themselves uniformly over the conductor surfaces. Consequently, the total field can be considered to be the sum of the two symmetrical radial fields drawn in Figs. 34*B* and *C*.

The field surrounding the charged conductor in Fig. 34*B* has the same configuration as the field surrounding the inner conductor of the coaxial line in Fig. 33. The voltage between the charged conductor and a point at the center of the uncharged conductor is therefore given by Eq. 75, provided r_2 is replaced by d_0 and r_1 by r_0. Because d_0 is much larger than r_0, this voltage may be taken as a good approximation to the voltage between conductors due to the field of the charge $-q$ acting alone. An equal voltage between conductors results from the charge $+q$, and because the fields due to the charges on the conductors are in the same direction, the total voltage e between conductors is twice either of the component voltages. Thus

$$e = \frac{q}{\pi\varepsilon} \ln \frac{d_0}{r_0} \tag{78}$$

and c, the capacitance per unit length of line, is

$$c = \frac{q}{e} = \frac{\pi\varepsilon}{\ln (d_0/r_0)} \tag{79}$$

For a line having a dielectric material of unit dielectric constant, Eq. 79 becomes

$$c = \frac{27.8}{\ln (d_0/r_0)} \qquad \mu\mu\text{f per m} \tag{80}$$

17. Transmission-line Inductance. In the determination of l, the inductance per unit length of a transmission line, a current is assumed to circulate in the line conductors, and the resulting magnetic flux that links the conductors is determined. The value of l is the ratio for a unit length of line of the magnetic flux to the current linked by the flux.

The magnetic-field quantities employed in the calculation are defined in a manner analogous to the definitions of the electric-field quantities. The magnetic field may be described by flux lines that represent the *magnetic-flux density*, a vector quantity directed along the flux lines and having a magnitude $\mathcal{B}$ determined by the density of the lines. The unit for $\mathcal{B}$ in the mks system is determined by reference to Faraday's law which states: *The electromotive force around a closed path in a magnetic field is proportional to the time rate of change of the magnetic flux that links the path.* In the mks system, the constant of proportionality is unity, and thus the unit for the total flux Φ is the volt-second. A special name, the *weber*, has been assigned to this unit, but it should be remembered that the weber is a volt-second, and thus a magnetic flux that changes at the rate of one weber per second gives rise to an electromotive force of one volt. Flux density $\mathcal{B}$ is measured in webers per square meter in the mks system.

The magnetic field, like the electric field, can be described alternatively in terms of equipotential surfaces. The *magnetic-field intensity* is a vector

having a direction perpendicular to the surfaces of constant magnetic potential and a magnitude $\mathcal{H}$ determined by the potential gradient, or rate of change of magnetic potential along a line perpendicular to the equipotential surfaces. The mks units of magnetic potential and magnetic-field intensity are determined by reference to Ampère's law: *The magnetomotive force around a closed path is proportional to the current enclosed by the path.* In the mks system, the proportionality constant is unity, and thus the unit of magnetic potential or magnetomotive force is the same as that of current, the ampere. Magnetic-field intensity is therefore measured in amperes per meter.

For magnetic fields in space or in any material having the same magnetic properties in all directions, the flux-density and field-intensity vectors are aligned at each point. Furthermore, except in ferromagnetic materials, the magnitudes of the two vectors have a constant ratio determined by the material. That is,

$$\mathcal{B} = \mu\mathcal{H} \tag{81}$$

where μ is a proportionality constant called the permeability of the material. In the mks system, the unit of μ, or $\mathcal{B}/\mathcal{H}$, is

$$\frac{\text{Webers/square meter}}{\text{Amperes/meter}} = \frac{\text{volt-seconds/ampere}}{\text{meter}} = \frac{\text{volts/(ampere/second)}}{\text{meter}} = \text{henries/meter}$$

The permeability of space is designated μ_0 and has the value

$$\mu_0 = 4\pi \times 10^{-7} \text{ henry per m} = 1.257 \ \mu\text{h per m}$$

For materials ordinarily employed in transmission lines, the permeability is negligibly different from μ_0.

In a coaxial line the magnetic-field pattern takes the form of concentric circles in the space between conductors, as indicated in Fig. 35*A*. In this figure a current i is assumed to flow in the conductors, the outer-conductor current being directed into the page and the inner-conductor current out of the page. A magnetic field encircles the current in the inner conductor, and the direction of the field is related to the direction of current flow by the right-hand rule: *A right-handed screw with its axis along a current-carrying conductor advances in the direction of current flow when rotated in the direction of the magnetic field surrounding the conductor.* The magnetic field beyond the outer conductor of the line is zero because the net current encircled by any path entirely surrounding both conductors is zero.

At the high frequencies employed in radar transmission lines, the current is confined almost entirely to the inside surface of the outer con-

ductor and to the outside surface of the inner conductor. The magnetic field is therefore restricted to the dielectric space between conductors. The confinement of current to the conductor surfaces is called skin effect, and it is the result of the magnetic field produced by the current. Skin effect is explained in detail in the next article in connection with the calculation of transmission-line resistance.

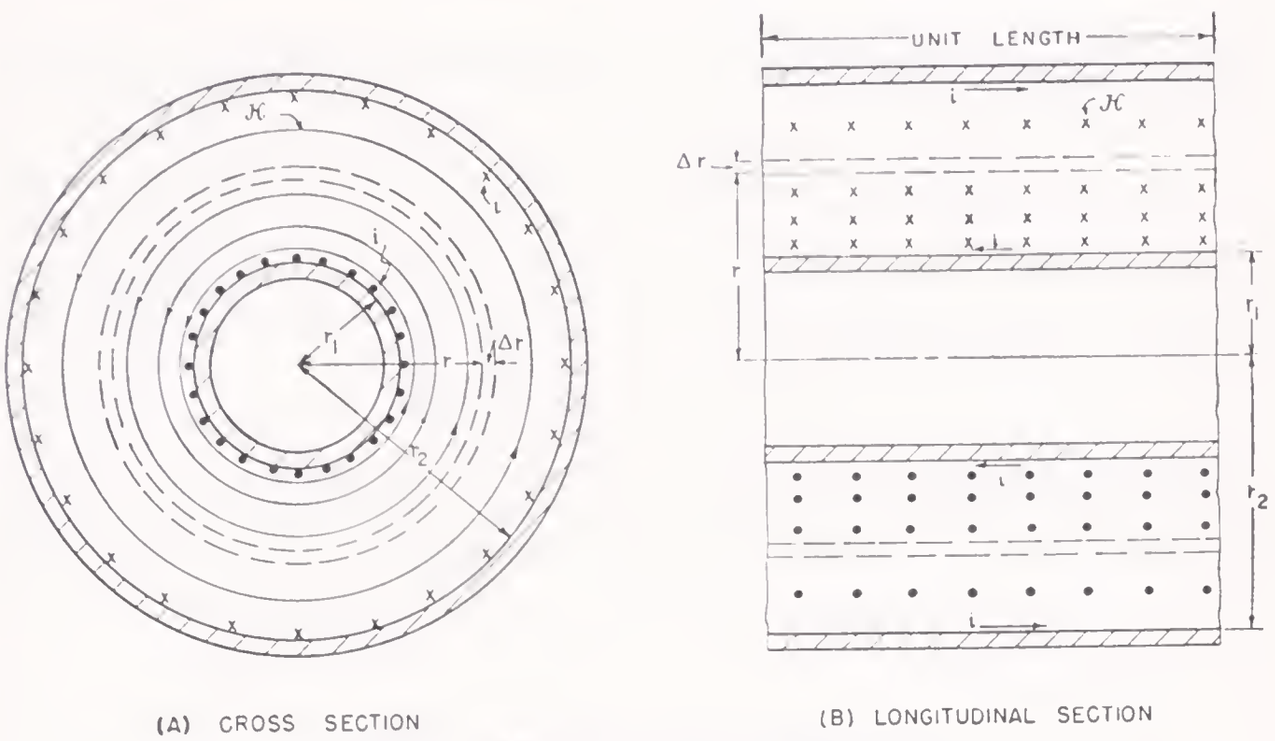


FIG. 35. Magnetic field in coaxial line.

The magnetomotive force around each circular flux path in Fig. 35*A* is equal to the encircled current i. Because of the symmetry of the line, the magnetic-field intensity has constant magnitude around the flux path, and therefore

$$\mathcal{H} = \frac{\text{magnetomotive force}}{\text{path length}} = \frac{i}{2\pi r} \tag{82}$$

where r is the radius of the flux circle. Thus the magnetic-field intensity of a coaxial line, like the electric-field intensity, varies inversely with the distance from the axis of the line. The magnetic-flux density $\mathcal{B}$ at any point between conductors is, by Eqs. 81 and 82,

$$\mathcal{B} = \mu\mathcal{H} = \frac{\mu i}{2\pi r} \tag{83}$$

The total flux linking the current in a line section of unit length is obtained by dividing the space between conductors into a series of cylindrical shells of thickness Δr (see Fig. 35*B*) and by adding the fluxes included within each shell. The flux in any shell is approximately the product of $\mathcal{B}$ by the cross-sectional area of the shell $1 \times \Delta r$, and the total

flux Φ is therefore given by the integral

$$\Phi = \int_{r_1}^{r_2} \mathcal{B}\, dr = \int_{r_1}^{r_2} \frac{\mu i}{2\pi r}\, dr = \frac{\mu i}{2\pi} \int_{r_1}^{r_2} \frac{dr}{r}$$

or

$$\Phi = \frac{\mu i}{2\pi} \ln \frac{r_2}{r_1} \tag{84}$$

Because this flux links the current i, the inductance of the unit-length line section is Φ/i, or

$$l = \frac{\Phi}{i} = \frac{\mu}{2\pi} \ln \frac{r_2}{r_1} \tag{85}$$

Because, for materials normally used as the dielectric in transmission lines, μ is very nearly equal to μ_0,

$$l = 0.2 \ln \frac{r_2}{r_1} \qquad \mu\text{h per m} \tag{86}$$

The magnetic-field pattern in the transverse plane of a parallel-wire line is drawn in Fig. 36. As in the coaxial line, the flux at high frequencies

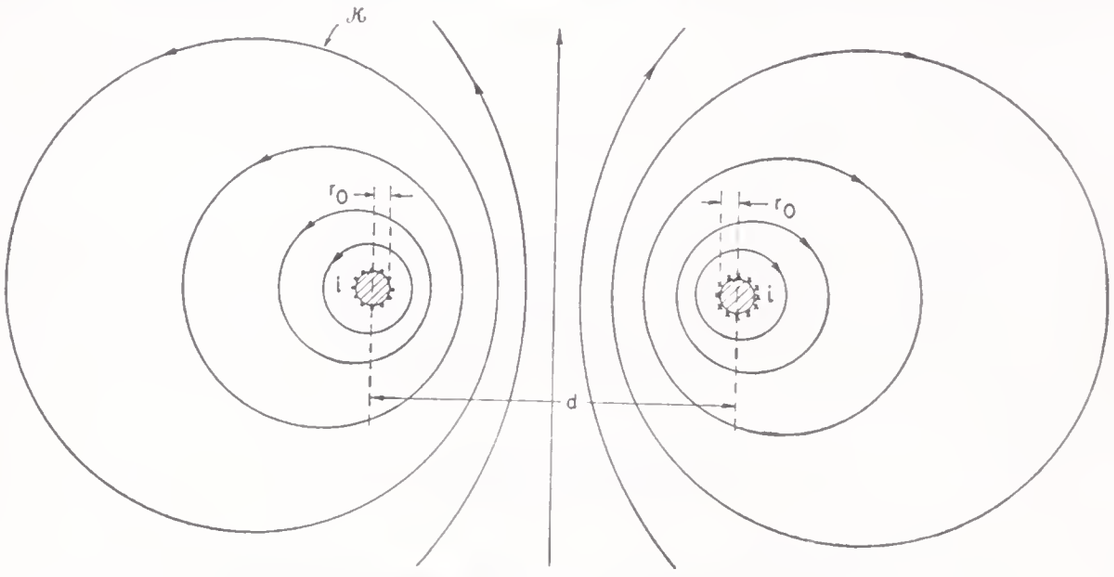


FIG. 36. Magnetic field of parallel-wire line.

may be considered to be entirely outside the conductors because of skin effect. The flux linking the line current is the flux between conductors and, for the condition that r_0 is much less than d_0, is the sum of the component fluxes between conductors set up by the two conductor currents acting separately. The component flux patterns consist of concentric circles about each conductor, and therefore the flux between wires due to a current i in one conductor may be obtained from Eq. 84 by replacing r_2/r_1 by d_0/r_0. The total flux Φ is twice the value thus obtained; that is,

$$\Phi = \frac{\mu i}{\pi} \ln \frac{d_0}{r_0}$$

The inductance per unit length of line is therefore

$$l = \frac{\Phi}{i} = \frac{\mu}{\pi} \ln \frac{d_0}{r_0} \tag{87}$$

or, if the permeability is that of space,

$$l = 0.4 \ln \frac{d_0}{r_0} \qquad \mu\text{h per m} \tag{88}$$

18. Transmission-line Resistance. The resistance of a conductor depends upon the resistivity of the material, the length and cross section of the conductor, and the distribution of the current in the cross section. In d-c circuits, current is uniformly distributed throughout the cross sections of conductors, and the resistance of a unit length of conductor of a given resistivity depends upon the cross-sectional area of the conductor. In a-c circuits, the changing magnetic flux associated with the current causes a nonuniform current distribution and increases the effective resistance of the conductor. This effect, known as *skin effect*, takes an extreme form at the high frequencies associated with radar transmission lines and forces the current to flow almost entirely in a thin layer near the surface of the conductor. Thus the high-frequency resistance of a conductor is determined by the distance around its cross section, rather than by the cross-sectional area.

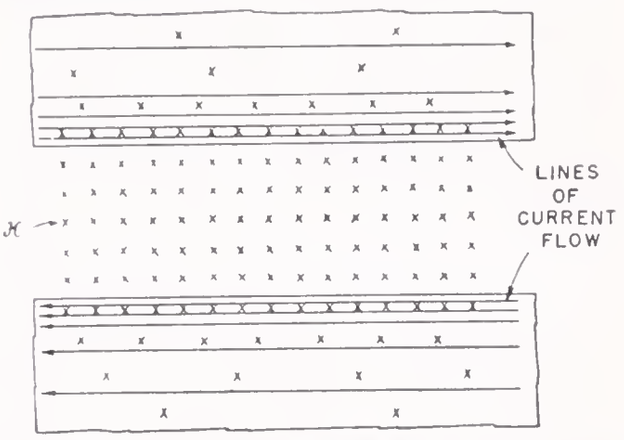


FIG. 37. Longitudinal section of transmission line showing current-flow lines and associated magnetic field.

Figure 37 shows part of a *longitudinal* section of a transmission line having closely spaced conductors. The arrows in the conductors indicate paths of current flow, and the crosses represent magnetic flux directed into the page. Because the current is not restricted completely to the surface, the magnetic field penetrates the conductors to some extent. Thus a current path far inside the conductor is linked by more flux than is a current path near the surface, and a greater inductance is associated with the internal path. Because of this variation of inductance with distance from the surface, an alternating current in the conductor becomes less in magnitude and more lagging in phase as the depth of the path increases. At low frequencies, variations in the current magnitude and phase over the conductor cross section are slight. However, increasing the frequency increases the reactance of the inductance and thus increases the effect until, at radar frequencies, the current is confined to a thin shell at the surface.

The current distribution in the conductor may be considered the result of a highly attenuated wave that propagates from the conductor surface into the body of the conductor. The amplitude of the current density decreases exponentially with distance beneath the surface as indicated by the solid curve in Fig. 38, and the phase changes linearly with distance. The distance δ, called the *skin depth*, is analogous to the time constant of an exponential function of time; that is, at a depth δ, the current density has decreased to a value which is $1/\epsilon$ times its surface value. The current at this depth lags the surface current by 1 radian.

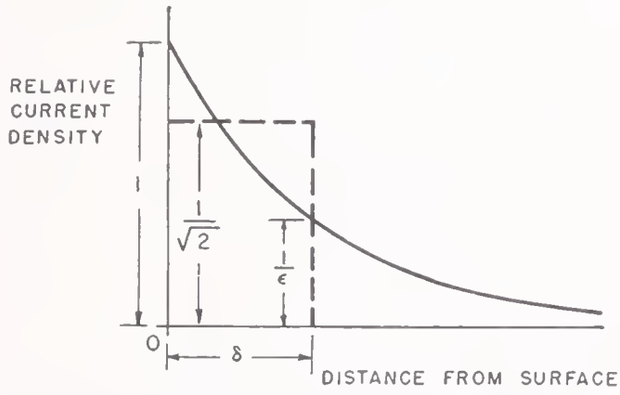


FIG. 38. Curves of current density as a function of distance from the surface of a conductor.

Another interpretation of δ is indicated by the dashed rectangular curve in Fig. 38. It can be shown[1] that the actual current-density distribution can be replaced by a fictitious current density of constant magnitude and phase in a surface layer of thickness δ with no change in the total current or total power loss, provided the fictitious current density is $1/\sqrt{2}$ times the actual current density at the surface. The equivalent current distribution is convenient for calculating transmission-line resistance.

The magnitude of the skin depth is given by the expression

$$\delta = \sqrt{\frac{\rho}{\pi \mu_c f}} \tag{89}$$

where f is the operating frequency, ρ is the resistivity, and μ_c the permeability of the conductor material (the subscript c is used to distinguish μ_c from the symbol μ employed for permeability of the dielectric in Art. 17). As an example of the magnitude of δ at radar frequencies, consider a copper conductor for which ρ is 1.72×10^{-8} ohm-meter and μ_c is very nearly the permeability of space, $4\pi \times 10^{-7}$ henry per m. Substitution of these values into Eq. 89 yields for the value of δ approximately 0.00026 in. at 100 Mcps and 0.00005 in. at 3,000 Mcps. The skin depth is therefore very small relative to normal conductor thickness, and the effective cross section is the conductor perimeter multiplied by the skin depth.

For a coaxial line of conductor radii r_1 and r_2, the effective cross section of the inner conductor is $2\pi r_1 \delta$ and of the outer conductor, $2\pi r_2 \delta$. For the parallel-wire line having two circular conductors each of radius r_0, the

[1] H. A. WHEELER, "Formulas for the Skin Effect," *Proc. I.R.E.*, **30** (September, 1942), 412–424.

effective cross section of each conductor is $2\pi r_0\delta$. Thus the resistance r of a unit length of coaxial line is

$$r = \frac{\rho}{2\pi r_1\delta} + \frac{\rho}{2\pi r_2\delta} = \frac{\rho}{2\pi\delta}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)$$

or

$$r = \frac{1}{2}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)\sqrt{\frac{\rho\mu_c f}{\pi}} \tag{90}$$

Similarly, for the parallel-wire line, the resistance per unit length is

$$r = \frac{\rho}{2\pi\delta}\left(\frac{1}{r_0} + \frac{1}{r_0}\right) = \frac{1}{r_0}\sqrt{\frac{\rho\mu_c f}{\pi}} \tag{91}$$

Note that the inner conductor of a coaxial line contributes the greater part of the resistance because of its smaller perimeter.

Because conductor surface areas are limited by physical considerations, low-resistivity materials are necessary for low-loss, high-frequency transmission lines. Copper and brass are the usual conductor materials, and brass conductors are often silver plated. Because of the small skin depth, even a very thin plating is effective. At high frequencies, a reduction in resistivity is less effective in reducing conductor resistance because it is accompanied by a decrease in skin depth. Thus at frequencies for which the depth of penetration is small, resistance varies as the square root of the resistivity; at low frequencies, where the entire cross section is effective, resistance varies directly with the resistivity.

19. Characteristic Resistance, Propagation Speed, and Attenuation. The parameters upon which the operating characteristics of a transmission line of small loss are directly dependent—the characteristic resistance R_c, the speed of wave propagation v, and the attenuation constant α—may be determined in terms of line dimensions and materials through use of the relations for capacitance, inductance, and resistance developed in the three preceding articles.

For a *coaxial line* the values of c and l of Eq. 76, p. 533, and Eq. 85, p. 538, apply. Thus the characteristic resistance is

$$R_c = \sqrt{\frac{l}{c}} = \frac{1}{2\pi}\sqrt{\frac{\mu}{\varepsilon}}\ln\frac{r_2}{r_1} \tag{92}$$

and the propagation speed is

$$v = \frac{1}{\sqrt{lc}} = \frac{1}{\sqrt{\mu\varepsilon}} \tag{93}$$

For a *parallel-wire line*, Eq. 79, p. 535, and Eq. 87, p. 539, yield values of

c and l. Therefore,

$$R_c = \frac{1}{\pi}\sqrt{\frac{\mu}{\varepsilon}}\ln\frac{d_0}{r_0} \tag{94}$$

and v has the value given by Eq. 93.

The attenuation constant α (see Eq. 60, p. 517) is $\frac{1}{2}(r/R_c + gR_c)$. The value for g has not been calculated in the preceding articles. However, in a special case of particular interest—that of lines using air or an insulating gas as the dielectric—the dielectric losses are so small relative to losses in the conductors that g may be taken to be zero. For *coaxial lines,* values of r from Eq. 90, p. 541, and R_c from Eq. 92 yield (for $g = 0$)

$$\alpha = \frac{1}{2}\sqrt{\pi\mu_c\frac{\varepsilon}{\mu}}\,\frac{1/r_1 + 1/r_2}{\ln\,(r_2/r_1)}\,\sqrt{\rho f} \tag{95}$$

For *parallel-wire lines,* Eq. 91 and Eq. 94 yield values of r and R_c, and (if $g = 0$)

$$\alpha = \frac{1}{2}\sqrt{\pi\mu_c\frac{\varepsilon}{\mu}}\,\frac{1/r_0}{r_0\ln\,(d_0/r_0)}\,\sqrt{\rho f} \tag{96}$$

These equations are particularly useful with transmission lines comprising rigid parallel-wire or coaxial conductors and an air or insulating-gas dielectric. Not only is the assumption $g = 0$ allowable, but also the empty-space values of permittivity and permeability may be employed. For μ and μ_c each equal to μ_0 and ε equal to ε_0, Eqs. 92 through 96 become, in mks units,

For the *coaxial line:*

$$R_c = 60\ln\frac{r_2}{r_1} \quad \text{ohms} \tag{97}$$

$$\alpha = 2.64\times10^{-6}\,\frac{1/r_1 + 1/r_2}{\ln\,(r_2/r_1)}\,\sqrt{\rho f} \quad \text{nepers per m} \tag{98}$$

For the *parallel-wire line:*

$$R_c = 120\ln\frac{d_0}{r_0} \quad \text{ohms} \tag{99}$$

$$\alpha = 2.64\times10^{-6}\,\frac{\sqrt{\rho f}}{r_0\ln\,(d_0/r_0)} \tag{100}$$

For the *coaxial* and *parallel-wire* lines:

$$v = 2.998\times10^{8} \text{ m per sec} \tag{101}$$

That the speed of waves on both coaxial and parallel-wire lines is the speed of light is one manifestation of a general principle further illustrated in the next chapter: All transverse electromagnetic waves (waves in which the electric and magnetic fields are both perpendicular to the direction of propagation) travel through empty space with the same speed, 2.998×10^8 m per sec.

As an example of the values obtained from Eqs. 97 through 100, consider a $1\frac{5}{8}$-in. rigid coaxial line having a $\frac{5}{8}$-in. inner-conductor and a 0.049-in. outer-conductor wall. Assume that the conductors are brass, the dielectric is air, and the operating frequency is 3,000 Mcps. The inside radius of the outer conductor is 0.764 in. (0.0194 m), the outside radius of the inner conductor is 0.313 in. (0.00794 m), and the resistivity of brass is 7.02×10^{-8} ohm-meter. The characteristic resistance of the line is, by Eq. 97,

$$R_c = 60 \ln \frac{0.0194}{0.00794} = 53.7 \text{ ohms}$$

and the attenuation constant, by Eq. 98, is

$$\alpha = 2.64 \times 10^{-6} \times \frac{1/0.0194 + 1/0.00794}{\ln (0.0194/0.00794)} \times \sqrt{1.72 \times 10^{-8} \times 3 \times 10^9}$$
$$= 0.00786 \text{ neper per m} = 0.0682 \text{ db per m}$$

Measured characteristics of rigid lines may differ slightly from the calculated values. Especially, the attenuation may be greater because of the effect of minor irregularities such as stub or bead supports, joints, and bends in the line.

The equations of this article are less useful with flexible coaxial cables using solid dielectrics than they are with rigid gas-insulated lines. One reason is that the value of the dielectric constant of the solid dielectric, which is considerably greater than unity, must be known before R_c or v can be calculated. Another reason is that the attenuation is considerably increased by dielectric losses and by the increased conductor losses caused by the braiding of the outer conductor. Because tables of the conductor and dielectric properties must in any event be utilized, it is convenient to tabulate the final results directly. Some physical and electrical properties of flexible cables ordinarily employed in radar are indicated in Table 1, the Army-Navy type numbers being used to designate the cables. Note that, because R_c equals $\sqrt{l/c}$ and v equals $1/\sqrt{lc}$,

$$v = \frac{1}{R_c c} \tag{102}$$

Therefore, the propagation speed can be determined from the values of capacitance and characteristic resistance listed in the table.

TABLE 1. PROPERTIES OF SOLID-DIELECTRIC COAXIAL CABLES

Physical properties			*Electrical properties*			
Type No.	*Nominal over-all diameter, in.*	*Outer braid*	*Nominal characteristic impedance, ohms*	*Nominal capacitance, $\mu\mu f/ft$*	*Attenuation at 10 cm, db/100 ft*	*Maximum operating voltage, rms kv*
RG-5/U	0.332	Double	53.5	28.5	20	2.0
RG-8/U	0.405	Single	52.0	29.5	18	4.0
RG-9/U	0.420	Double	51.0	30.0	18	4.0
RG-10/U*	0.475	Single	52.0	29.5	18	4.0
RG-14/U	0.545	Double	52.0	29.5	14	5.5
RG-17/U	0.870	Single	52.0	29.5	9.6	11.0
RG-18/U*	0.945	Single	52.0	29.5	9.6	11.0
RG-19/U	1.120	Single	52.0	29.5	8.3	14.0
RG-20/U*	1.195	Single	52.0	29.5	8.3	14.0
RG-55/U	0.206	Double	53.5	28.5	34	1.9
RG-58/U	0.195	Single	53.5	28.5	34	1.9
RG-74/U	0.615	Double	52.0	29.5	13.8	5.5

* Armored.

CHAPTER VIII

WAVEGUIDES

In the preceding chapter the two-wire transmission line was discussed as a means of guiding radio-frequency power from one point to another. In this chapter it will be shown that two conductors are not always necessary in order to transfer radio-frequency power between two points in space. A single conductor, usually in the form of a metal pipe or *waveguide*, is sufficient if the frequency is high enough, and in the microwave region the waveguide is often preferable to the two-wire line. Even the single conductor is not essential for the transfer of radio-frequency power; a dielectric rod can serve to guide waves from one point to another. However, the dielectric rod has the disadvantage of lower efficiency than the metallic conductor, and except in a few very specialized applications, only the metallic conductor is of practical importance.

A two-wire transmission line may be analyzed conveniently in terms of the ordinary electric-circuit quantities of voltage, current, and impedances. However, in a description of waveguide operation, the electric-circuit concepts are of little use because of the complicated manner in which the voltage and current are distributed in the guide. A more convenient approach makes use of the *electric* and *magnetic fields* within the guide.

1. Applications and Advantages of Waveguides. Waveguides perform the same functions in radar as do two-conductor transmission lines; that is, they serve to transfer radio-frequency power from the transmitter to the antenna during the transmission of pulses and from the antenna to the receiver during the reception of echoes. Waveguides are used principally at frequencies in the microwave region; inconveniently large guides would be required to transmit radio-frequency power at the longer wavelengths.

It is possible to propagate within a waveguide several different types of electromagnetic waves, each of which is characterized by a different electric- and magnetic-field configuration. Associated with each wave type is a *cutoff frequency* below which, for a given size guide, propagation becomes impossible. In general, the larger the guide, the lower is the cutoff frequency of each type of wave.

The wave with the lowest cutoff frequency is called the *dominant* wave. Normally, radar systems use rectangular waveguides with dimensions

chosen so that only the dominant wave is propagated. For a rectangular guide a simple relation exists between the cutoff frequency of the dominant wave and the inside width of the guide. If the inside width is denoted by a, as in Fig. 1, and if c is the speed of electromagnetic waves in space (3×10^8 m per sec), the cutoff frequency f_{co} is $c/2a$. Since the corresponding free-space wavelength λ_{co} is c/f_{co},

$$\lambda_{co} = 2a \tag{1}$$

Thus, to transmit the dominant wave, a rectangular guide must have a width of at least one-half the free-space length of the wave. The b, or shorter cross-sectional, dimension is not critical; usually it is made about one-half the longer dimension.

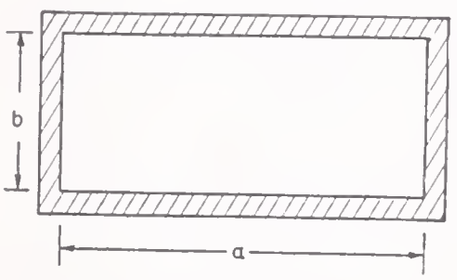


FIG. 1. Cross section of a rectangular waveguide showing a and b dimensions.

To ensure that only the dominant wave is propagated, the cross-sectional dimensions of the guide must be maintained near the minimum allowable values. Higher order waves may be propagated if the a dimension exceeds λ_0, the free-space wavelength, or if the b dimension exceeds $\lambda_0/2$. In practice a is usually about $0.7\lambda_0$ and b about $0.35\lambda_0$. To transmit a 10-cm wave, for example, a rectangular waveguide must have a width greater than 5 cm and in practice is made about 7 cm wide. A typical section has outside dimensions of 3 by $1\frac{1}{2}$ in. For a 3-cm wave, a smaller cross section is used. A standard size is 1 by $\frac{1}{2}$ in.

Where a circularly symmetrical field pattern is required in a waveguide transmission system, as, for example, at a rotary joint (Art. 3, Chap. XII), a section of circular waveguide is substituted for the rectangular guide. The 10-cm circular guides used for this purpose have diameters of approximately 3 in., and for similar 3-cm guides the diameter is about 1 in.

Waveguide dimensions for operation at 10 cm and below are therefore entirely reasonable, but for longer waves the size required is impractical. The equivalent of the 3-in. rectangular guide for 10-cm waves would have to be 30 in. (2.5 ft) wide if the wavelength were 1 m and 25 ft if the wavelength were 10 m.

In waveguides the electric and magnetic fields are confined to the space within the guides. Hence, no power is lost through radiation, nor is dielectric loss of any practical importance, since the guides are normally air-filled. With respect to radiation and dielectric loss, therefore, air-filled coaxial lines and waveguides are equally efficient. The over-all attenuation of a waveguide is lower, however, because at a given frequency the power lost as heat in the walls of a standard-size waveguide is less than the power dissipated as heat in the conductors of conventional-size coaxial lines. For example, at 10 cm the attenuation of a standard-

size (1⅝-in. outer conductor, ⅝-in. inner conductor), rigid, stub-supported coaxial line made of brass is nearly 0.07 db per m (see Art. 19, Chap. VII), whereas a 3- by 1-in. waveguide of the same material has an attenuation of only 0.04 db per m. Increasing the size of the coaxial line to decrease the attenuation is not permissible because of the possibility of transmitting higher order types of waves.

Another advantage of waveguides is that their peak power capacity is greater than that of coaxial lines. This advantage is especially important at 3 cm, where the size of the coaxial line must be kept small in order to ensure transmission of only the principal wave. The maximum peak power that can be transmitted by a ½-in. stub-supported line at 3 cm is about 360 kw. The corresponding power for a 1- by ½-in. waveguide is 1 Mw.

Waveguides have the further advantage of being simple in construction because no inner conductor or conductor supports are needed. They are also more rugged than coaxial lines, the inner conductor of which may become displaced or even damaged by vibration or shock.

Either coaxial lines or rectangular waveguides may be employed in 10-cm radar systems. The coaxial line is preferable because of its compactness if the low attenuation or high-power capacity of the waveguide is not required. On the other hand, the waveguide is used if the distance from the antenna to the rest of the radar set is so great that low attenuation is important or if peak power of the order of several megawatts must be transmitted. At 3 and 1 cm, waveguides are used almost exclusively. At these wavelengths, waveguides are compact, their attenuation is lower than that of coaxial lines, and their power capacity is higher.

2. Basic Electromagnetic-wave Concepts—The Parallel-strip Transmission Line. In preparation for the study of waveguides, it is desirable to consider the relation of waves of electric- and magnetic-field intensities to waves of voltage and current on transmission lines. A special form of line, the parallel-strip line pictured in Fig. 2*A*, is useful for this purpose because of its simple geometry.

The line of Fig. 2*A* comprises two parallel flat-plate conductors of width w and spacing h. For simplicity, the line is considered lossless; the conductors have zero resistance, and the dielectric has zero conductance. Incident and reflected waves of voltage and current may be propagated along the line in the manner described in Chap. VII. Consider that sinusoidal waves are traveling from left to right in Fig. 2*A*, and let the voltage between conductors and the current in either conductor at any cross section be denoted by e and i, respectively.

The two conductors are equivalent to two plates of a capacitor. Charges appear on the inner surfaces of the conductors, and an electric field is present between conductors because of the voltage e. The charge

and the field at any cross section increase and decrease in phase with the voltage at the section, and the pattern of the field is indicated by the solid-line arrows in Fig. 2*B*. If w is much larger than h, the electric field is uniform, except for a fringing effect near the edges of the strips. As explained in Art. 16 of Chap. VII, the intensity of the uniform electric field is

$$\mathcal{E} = \frac{e}{h} \tag{2}$$

where, in rationalized mks units, e is in volts, h in meters, and $\mathcal{E}$ in volts per meter. Algebraic signs are chosen in Eq. 2 on the basis that, when the voltage of the upper conductor is positive, the electric field, which is directed downward, is also positive.

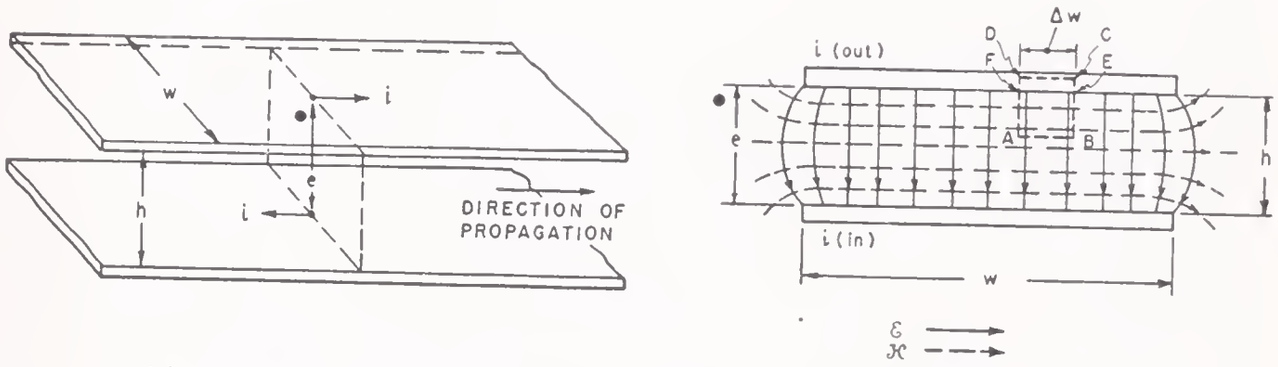


FIG. 2. Parallel-strip transmission line.

Current in the conductors causes a magnetic field to encircle each conductor. The direction of the field is related to the direction of the current by the right-hand rule (p. 536), as indicated by the dashed-line arrows of Fig. 2*B*. Because of the plane geometry of the conductors, the magnetic field is uniform in most of the region between conductors, as shown. Because the conductors are assumed to have no resistance, the skin depth (p. 540) is zero, and the current flows entirely in a sheet of zero thickness at the surface of the conductors. The magnetic field, as well as the electric field, is thus prevented from entering the conductors.

The relation of $\mathcal{H}$, the intensity of the magnetic field, to the current in the conductors may be determined by means of Ampère's law (p. 536). In Fig. 2*B*, the rectangular loop $ABCD$ encircles the portion of the sheet of surface current between points F and E. If $\mathcal{J}$ stands for the *linear* density of the surface current, that is, the current per unit width of the path of flow, and Δw is the width of the rectangular loop, then $\mathcal{J}\Delta w$ is the encircled current. Also, since sides BC, CD, and DA of the loop lie either in a region of zero field or perpendicular to the magnetic field, the magnetomotive force around the loop is entirely due to the field along side AB and is $\mathcal{H}\Delta w$. Thus Ampère's law states that $\mathcal{H}$ equals the linear density of the surface current. This result is true in general: At the boundary of

a perfect conductor and a perfect dielectric, a magnetic field parallel to the boundary is always associated with a current at the surface of the conductor. The field is perpendicular to the direction of flow of the current and equal in magnitude to the linear density of the current.

In Fig. 2*B*, if w is much larger than h, the uniform magnetic-field intensity between conductors is much stronger than the field intensity in the space outside the conductors. Correspondingly, the current is restricted almost entirely to the inner surface of the conductors, so that for all practical purposes $\mathcal{J} = i/w$. Thus

$$\mathfrak{IC} = \mathcal{J} = \frac{i}{w} \tag{3}$$

where, in rationalized mks units, i is in amperes, w in meters, and $\mathfrak{IC}$ and $\mathcal{J}$ in amperes per meter. The positive reference direction of i is outward from the page in the upper conductor in Fig. 2*B*, and the positive reference direction of $\mathfrak{IC}$ is to the right.

The voltage and current distribution along a parallel-strip line at a single instant is illustrated in Fig. 3*B*. Equations 2 and 3 show that at each point along the line the electric-field intensity is proportional to the voltage and the magnetic-field intensity is proportional to the current. Thus, if the voltage and current waves are of the form

$$e = E_m \cos(\omega t - \beta_0 s) \tag{4}$$

and

$$i = I_m \cos(\omega t - \beta_0 s) \tag{5}$$

corresponding waves of electric- and magnetic-field intensity may be expressed as

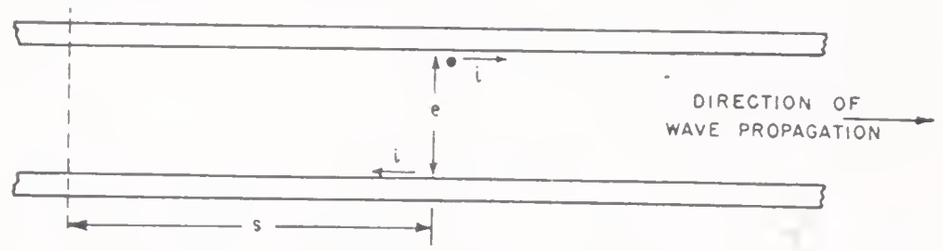
$$\mathcal{E} = \frac{e}{h} = \frac{E_m}{h} \cos(\omega t - \beta_0 s) = \mathcal{E}_m \cos(\omega t - \beta_0 s) \tag{6}$$

$$\mathfrak{IC} = \frac{i}{w} = \frac{I_m}{w} \cos(\omega t - \beta_0 s) = \mathfrak{IC}_m \cos(\omega t - \beta_0 s) \tag{7}$$

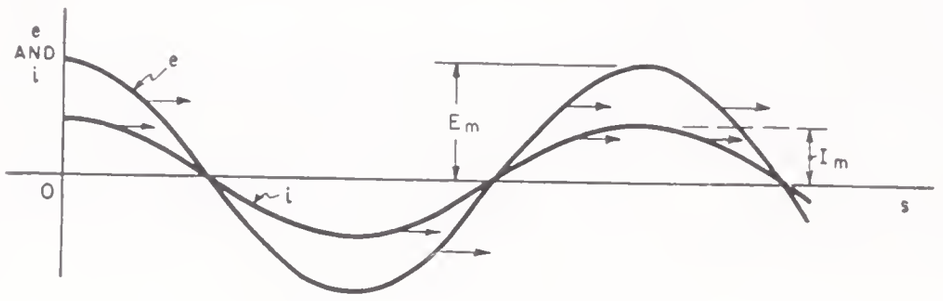
In these equations, E_m and I_m are the peak values of the voltage and current waves, $\mathcal{E}_m$ and $\mathfrak{IC}_m$ are the peak values of the electric- and magnetic-field-intensity waves, ω and β_0 are, respectively, the angular frequency and phase constant of the waves, and s is the distance indicated in Fig. 3*A*. Electric- and magnetic-field-intensity waves are drawn in Fig. 3*C* for the instant for which the voltage and current waves of Fig. 3*B* apply. Note that the variations in $\mathcal{E}$ and $\mathfrak{IC}$ along the line are identical with those of e and i except for scale factors.

The electric-field and magnetic-field waves described in Fig. 3*C* and Eqs. 6 and 7 are represented more pictorially in Figs. 4*A* and *B*. The diagrams of Fig. 4*A* make use of the uniformity of the $\mathcal{E}$ and $\mathfrak{IC}$ fields in a

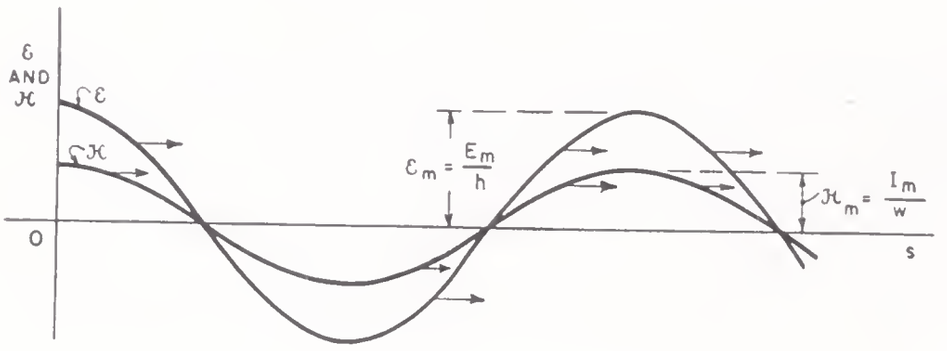
direction perpendicular to the conductors and show the magnitude of $\mathcal{E}$ and $\mathcal{H}$ as a function of position in a plane parallel to the conductors, the distance between the curved surface and reference plane in the drawing being proportional to the quantity represented. The diagrams of Fig. 4*B* show both magnitude and direction of the fields in three perpendicular sections of the line. The usual scheme of indicating direction of vectors by directions of lines and magnitude of vectors by closeness of spacing of



(A) CIRCUIT DIAGRAM



(B) VOLTAGE AND CURRENT WAVES



(C) ELECTRIC-FIELD AND MAGNETIC-FIELD WAVES

FIG. 3. Waves on a parallel-strip transmission line.

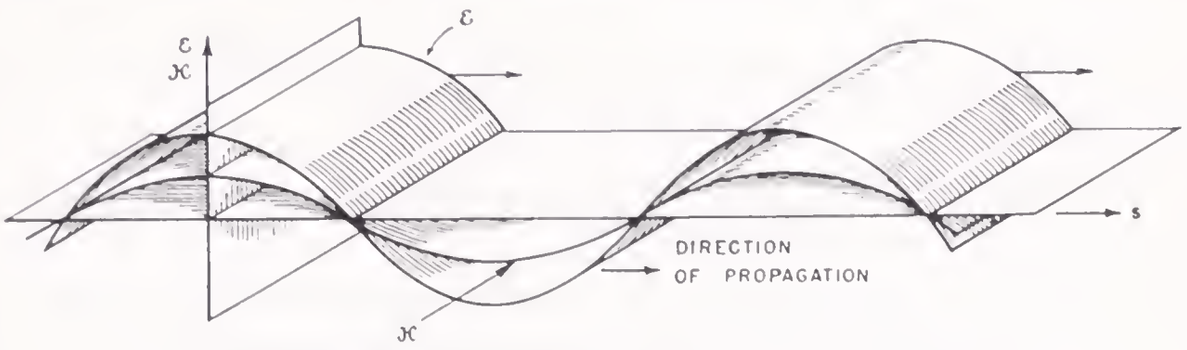
the lines is employed, solid lines being used for the electric-field vectors and dashed lines for the magnetic-field vectors.

An important parameter associated with the voltage and current waves is the characteristic resistance of the line R_c, which is the constant ratio that exists between the voltage and current at any cross section of the line for any traveling wave. An equally important parameter associated with the field-quantity waves is the *characteristic wave impedance* Z_c, which is defined as the ratio of $\mathcal{E}$ to $\mathcal{H}$ in any cross section of the line. Because of the relations of Eqs. 2 and 3, $\mathcal{E}$ and $\mathcal{H}$ have the constant ratio

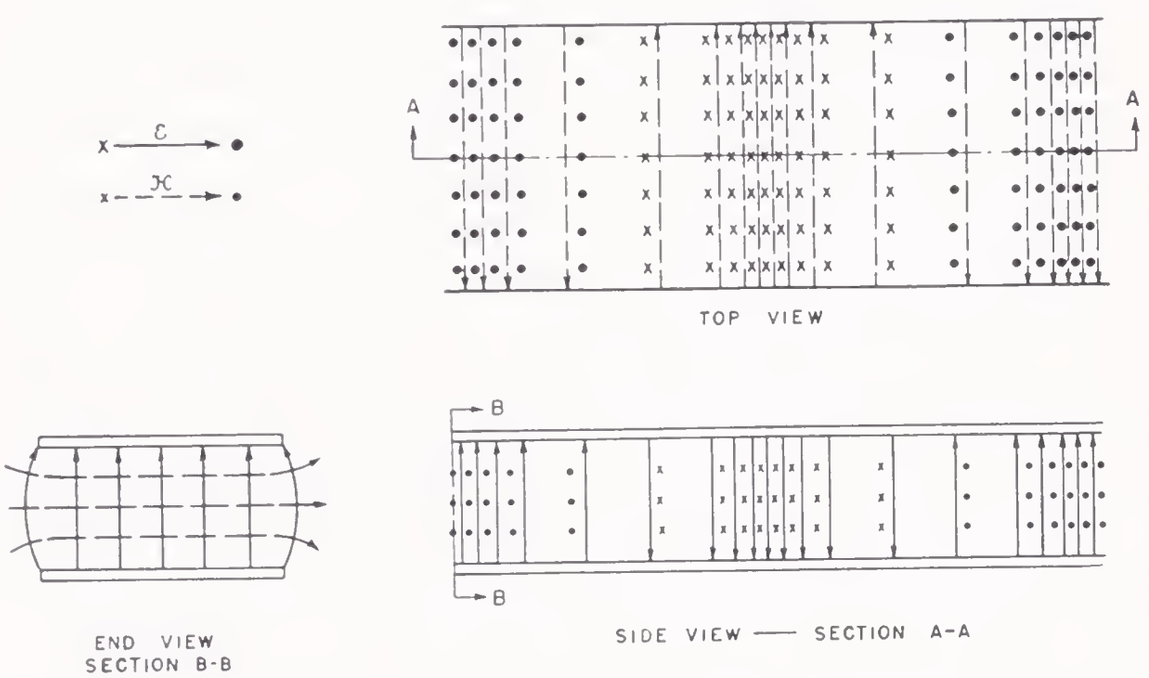
$$Z_c = \frac{\mathcal{E}}{\mathcal{H}} = \frac{e/h}{i/w} = \frac{w}{h} R_c = \frac{w}{h} \sqrt{\frac{l}{c}} \tag{8}$$

where l and c represent, respectively, the inductance per unit length and capacitance per unit length of the line.

Because of the simple geometry of the parallel-strip line, l and c are easily calculated. The capacitance c of a unit length of line is the



(A) TWO-DIMENSIONAL PLOT OF FIELD-QUANTITY MAGNITUDES



(B) SECTIONAL DRAWING OF FIELD SHOWING MAGNITUDE AND DIRECTION OF FIELD-QUANTITY VECTORS

FIG. 4. Representations of electric and magnetic fields in parallel-strip lines.

capacitance between two plates of area $w \times l$ units and separated a distance h; therefore,

$$c = \varepsilon \frac{w}{h} \tag{9}$$

where ε is the permittivity of the dielectric medium between plates. Similarly, the inductance l results from a magnetic flux Φ linking the conductors of a unit-length section of line. The flux equals the flux

density $\mu\mathfrak{IC}$ multiplied by the area $h \times l$, and

$$l = \frac{\Phi}{i} = \frac{\mu\mathfrak{IC}h}{i} = \frac{\mu(i/w)h}{i} = \mu\frac{h}{w} \tag{10}$$

In this expression, μ is the permeability of the dielectric medium.

Substitution of the values of l and c into Eq. 8 yields for $\mathcal{Z}_c$ the value

$$\mathcal{Z}_c = \frac{w}{h}\sqrt{\left(\frac{h}{w}\right)^2\left(\frac{\mu}{\varepsilon}\right)} = \sqrt{\frac{\mu}{\varepsilon}} \tag{11}$$

In rationalized mks units, $\mathcal{Z}_c$ is the ratio of volts per meter to amperes per meter and is thus measured in ohms.

Characteristic wave impedance may be considered an analogue of characteristic resistance in the same sense that $\mathfrak{E}$ is an analogue of e and $\mathfrak{IC}$ an analogue of i. Because the time variations of $\mathfrak{E}$ and $\mathfrak{IC}$ in the traveling wave of Fig. 4, like those of e and i, are in phase at each cross section of line, $\mathcal{Z}_c$ is the analogue of a purely resistive impedance. It must be remembered that $\mathcal{Z}_c$, though measured in ohms, is the ratio of electric-field intensity to magnetic-field intensity and is therefore *not* an impedance in the usual circuit sense of a voltage-to-current ratio. In the text, the term *wave impedance* and the script letter $\mathcal{Z}$ are used to designate ratios of $\mathfrak{E}$ to $\mathfrak{IC}$. The unmodified word *impedance* and the italic letter Z are reserved for the ratio of the circuit quantities e and i.

The need for distinguishing between impedance and wave impedance is illustrated by the following example: All parallel-strip transmission lines employing the same dielectric material have the same characteristic wave impedance $\mathcal{Z}_c = \sqrt{\mu/\varepsilon}$. This fact does not mean that any two of these lines can be joined without a mismatch occurring at the junction. To avoid mismatch, the line dimensions must be chosen so that the characteristic resistance is the same for the two lines joined.

It is significant that $\mathcal{Z}_c$ for the waves of Fig. 4 (and, in fact, for all transverse waves, see Art. 3) depends only upon the constants μ and ε of the dielectric material through which the waves travel. The ratio $\sqrt{\mu/\varepsilon}$ is called the *intrinsic wave impedance* of the dielectric and is denoted by $\mathcal{Z}_i$. For the waves of Fig. 4, the characteristic wave impedance, which is a property of the wave, equals the intrinsic impedance, which is a property of the dielectric. Other types of waves are important in waveguides for which this statement is not true (see Art. 13). For air as a dielectric medium, μ and ε have essentially the same values as for space; that is, $\mu = \mu_0 = 4\pi \times 10^{-7}$ henry per m, and

$$\varepsilon = \varepsilon_0 = 8.854 \times 10^{-12} \text{ farad per m}$$

The intrinsic impedance of air or space is

$$Z_i = \sqrt{\frac{\mu_0}{\varepsilon_0}} = \sqrt{\frac{4\pi \times 10^{-7}}{8.854 \times 10^{-12}}} = 377 \text{ ohms} \tag{12}$$

Another important parameter associated with traveling waves is the speed of propagation. For the voltage and current waves and therefore for the field-intensity waves, this speed is $1/\sqrt{lc}$, where l and c, the per-unit-length inductance and capacitance of the line, are given by Eqs. 9 and 10. The speed of propagation v for air-dielectric lines is therefore

$$v = \frac{1}{\sqrt{lc}} = \frac{1}{\sqrt{(\mu_0 h/w)(\varepsilon_0 w/h)}} = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 2.998 \times 10^8 \text{ m per sec} \tag{13}$$

and equals the speed of light in space. This same result was obtained for waves on parallel-wire lines and in coaxial lines in Art. 19 of Chap. VII. For transverse waves on uniform two-conductor lines having conductor cross sections of any shape, this same speed of propagation applies.

The product of e and i, the instantaneous voltage and current at any point of the parallel-strip line, is p, the instantaneous power transferred at the point of the line. Just as e and i have their field-quantity analogues in $\mathcal{E}$ and $\mathcal{H}$, so also does p have a field-quantity analogue, the *Poynting vector* $\mathcal{S}$. By definition, the Poynting vector has a magnitude equal to $\mathcal{E}\mathcal{H}$, the product of the instantaneous values of electric- and magnetic-field intensities, and is perpendicular to the vectors representing the electric and magnetic fields. The direction is determined by means of the right-hand rule—a right-handed screw with its axis perpendicular to the electric and magnetic field would advance in the direction of $\mathcal{S}$ if rotated in the direction in which the electric-field vector would turn in moving *toward* the magnetic-field vector. If the Poynting vector were drawn on the diagram of Fig. 4*B*, it would have the same direction at all points of the line, the direction being that in which the waves are traveling. The magnitude of the vector over any cross section would be constant, but because $\mathcal{E}$ and $\mathcal{H}$ vary with time and distance along the line, the magnitude of $\mathcal{S}$ would likewise vary with these quantities.

The Poynting vector specifies the *power per unit area* transmitted through a surface perpendicular to the vector. Since the magnitude of $\mathcal{S}$ is constant throughout a cross section of the parallel-strip line, the rate of energy flow through each unit cross-sectional area is $\mathcal{E}\mathcal{H}$, and the total instantaneous power transmitted through the entire cross section is $\mathcal{E}\mathcal{H} \times hw = \mathcal{E}h \times \mathcal{H}w$. The quantity $\mathcal{E}h \times \mathcal{H}w$ equals $e \times i$, so that the total instantaneous powers determined through use of field and circuit quantities are in agreement. The total *average* power transmitted, in terms of circuit quantities, is the product of the rms values of the voltage

and current waves, or $(E_m/\sqrt{2})(I_m/\sqrt{2}) = \frac{1}{2}E_m I_m$. Similarly, the average power per unit of cross-sectional area transmitted by the field waves is $\frac{1}{2}\mathcal{E}_m\mathcal{H}_m$.

The presence of waves of electric- and magnetic-field intensity between conductors of the parallel-strip line suggests a concept of transmission-line operation somewhat different from the circuit concept presented in Chap. VII. If emphasis is placed upon the electric and magnetic fields which surround the conductors, power may be considered to be transmitted, not because current flows in conductors between which a voltage is present, but because electromagnetic waves move through space. Interest is therefore focused upon the space between conductors in which the fields are located. Conducting surfaces play the less significant but still important role of guiding the waves to their destination and of providing impenetrable boundaries through which the waves cannot pass.

3. Types of Electromagnetic Waves. The waves in the parallel-strip transmission line are only one simple example of the many forms that electromagnetic waves may take. When the conducting boundaries have a less simple shape, a great variety of more complex field patterns is possible. The waves of the parallel-strip line may be described as *plane, transverse, linearly polarized*, electromagnetic waves. A discussion of these terms will serve to illustrate the various types of waves that are possible.

Plane Waves. Electromagnetic waves may be classified in accordance with the shape of their *equiphase surfaces*, or *wavefronts*. An equiphase surface is the locus of points at which the fields are in the same phase of their cycle of variation. The direction of propagation is perpendicular to the equiphase surfaces and consequently varies from point to point if the surface is not plane. Lines marking the direction of propagation in various parts of the wave are called *rays*. In Fig. 4*B*, the equiphase surfaces are plane and transverse to the longitudinal dimension of the line, and the rays are all parallel horizontal lines directed from left to right.

Plane waves can be produced by a source spread over a plane of infinite extent. The wavefronts are then parallel to the plane of the source. Other types of waves include *cylindrical* waves, which spread out radially in two dimensions from a line of sources of infinite length, and spherical waves, which radiate radially in three dimensions from a point source.

Transverse Waves. A plane tangent to an equiphase surface and therefore perpendicular to the direction of propagation is the *transverse* plane for a point in the wave. A wave for which both the electric- and magnetic-field vectors lie in the transverse plane is a *transverse-electromagnetic wave* (often abbreviated *TEM* wave). If only one of the field vectors lies in the transverse plane, the wave is either a *transverse-electric* (*TE*) or *transverse-magnetic* (*TM*) wave. Waves in which both vectors have com-

ponents outside the transverse plane exist, but it is always possible to separate these waves into component *TE* and *TM* waves and to consider the components separately.

Both field vectors for the wave of Fig. 4*B* are perpendicular to the direction of propagation. The wave is therefore completely transverse. Completely transverse waves cannot be propagated within a waveguide, but both transverse-electric and transverse-magnetic waves are possible.

Linearly Polarized Waves. If the components of the electric and magnetic fields in the transverse plane do not change in direction from instant to instant or from point to point, the wave is *linearly polarized.* The transverse wave of Fig. 4*B* is linearly polarized because, as the wave moves between the conducting plates, the electric-field vector is always perpendicular to the conductors and the magnetic-field vector remains parallel to the conductors. The *direction* of *polarization* is taken as the direction of the line along which the electric-field vector lies. Thus, if the conductors of Fig. 4*B* lie in horizontal planes, the electric vector is in the vertical direction, and the waves are vertically polarized. Frequently, a useful description of a wave may be given by plotting the shape of an equiphase surface and marking the direction of polarization of the wave at various points of the surface.

If vertically and horizontally polarized transverse waves are combined, and if the phases of the variations of the electric-field intensity in the two waves are different, the direction of the resultant electric field will vary with time and distance. In any transverse plane, the tip of the resultant electric-field vector follows an elliptical path, the ellipse becoming a circle if the component waves are of equal amplitude and 90 deg out of phase. The resultant wave is said to be *elliptically polarized* or, in the special case, *circularly polarized.* An elliptically polarized wave may, for convenience in study, be separated into component linearly polarized waves.

Study of the parallel-strip transmission line in Art. 2 has shown that, for air as the dielectric, the plane, transverse, linearly polarized waves which travel between the strips have the following properties:

1. The ratio of electric-field intensity to magnetic-field intensity at any point of the traveling wave is constant and equal to the intrinsic wave impedance of space, 377 ohms.

2. The waves are propagated with the speed of light, 2.998×10^8 m per sec.

3. At every point of the wave, the Poynting vector has the same direction as a ray of the wave.

These properties are characteristic of all *completely transverse* waves propagating in air or in space, regardless of the shape of the equiphase surfaces or the guiding conductors—in fact, the guiding conductors may be entirely absent. If transverse waves are propagated in a dielectric

medium for which μ or ε has a value different from the values for space, different values of $\mathcal{Z}_i$ and propagation speed result. If the guiding conductors prevent the wave from being completely transverse, the speed of propagation, the ratio of electric-field intensity to magnetic-field intensity, and the direction of the Poynting vector are altered.

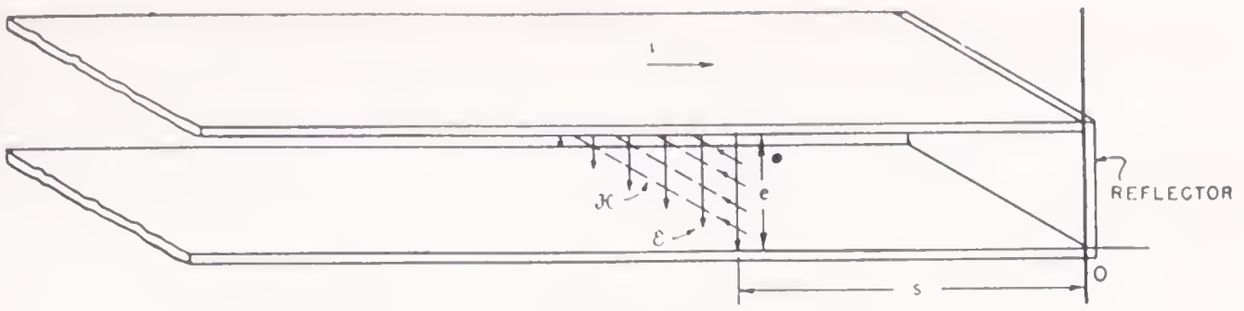
4. Normal Reflection of Plane Transverse Waves. Figure 5*A* shows a parallel-strip line terminated by a short circuit. The short-circuiting device consists of a perfectly conducting surface placed perpendicularly to the line conductors and to the direction of wave propagation. As indicated in Art. 3, Chap. VII, incident waves of voltage and current, upon striking the short circuit, give rise to reflected waves of voltage and current, and the traveling waves combine to produce the standing waves of voltage and current drawn in Fig. 5*B*.

Accompanying the incident voltage wave e^+ is an incident electric-field wave of intensity $\mathcal{E}^+ = e^+/h$, and accompanying the incident-current wave i^+ is an incident magnetic-field wave of intensity $\mathcal{H}^+ = i^+/w$. Similarly, associated with the reflected voltage and current waves e^- and i^- are reflected electric- and magnetic-field waves of intensities $\mathcal{E}^- = e^-/h$ and $\mathcal{H}^- = i^-/w$, respectively. Reference directions for the fields are shown by the arrows in Fig. 5*A* and are assumed to be the same for both the incident and reflected waves. The traveling field waves combine to produce standing waves in the same manner as do the corresponding waves of voltage and current. Since voltage is proportional to electric-field intensity and current is proportional to magnetic-field intensity, the diagrams of Fig. 5*B* can also be considered to represent the standing waves of $\mathcal{E}$ and $\mathcal{H}$.

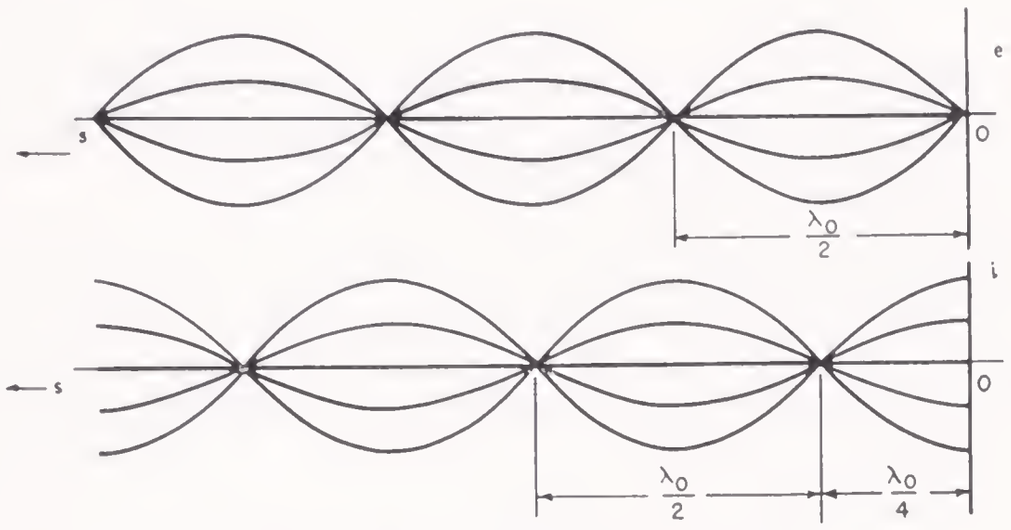
A pictorial representation of the electric-field-intensity waves is given in Fig. 5*C*. In this figure, the distance between the horizontal reference plane and the curved surface is proportional to the electric-field intensity at a particular point in a plane parallel to the transmission-line conductors. The upper two diagrams show the electric-field-intensity distribution due to the incident and reflected waves at one instant, and the lower diagram illustrates the resultant field intensity at several instants.

Because the voltage is always zero at the short-circuiting plane, the electric-field intensity is also zero in that plane. To fulfill this requirement, the values of $\mathcal{E}^+$ and $\mathcal{E}^-$ must always be equal in magnitude and opposite in direction *at the reflector*. The standing-wave pattern illustrated in the lower diagram of Fig. 5*C* results from this boundary condition in exactly the same way that the standing-wave pattern of voltage results from the requirement that $e^+ = e^-$ at the reflector; that is, for s measured from the reflector as in Fig. 5*A*, if

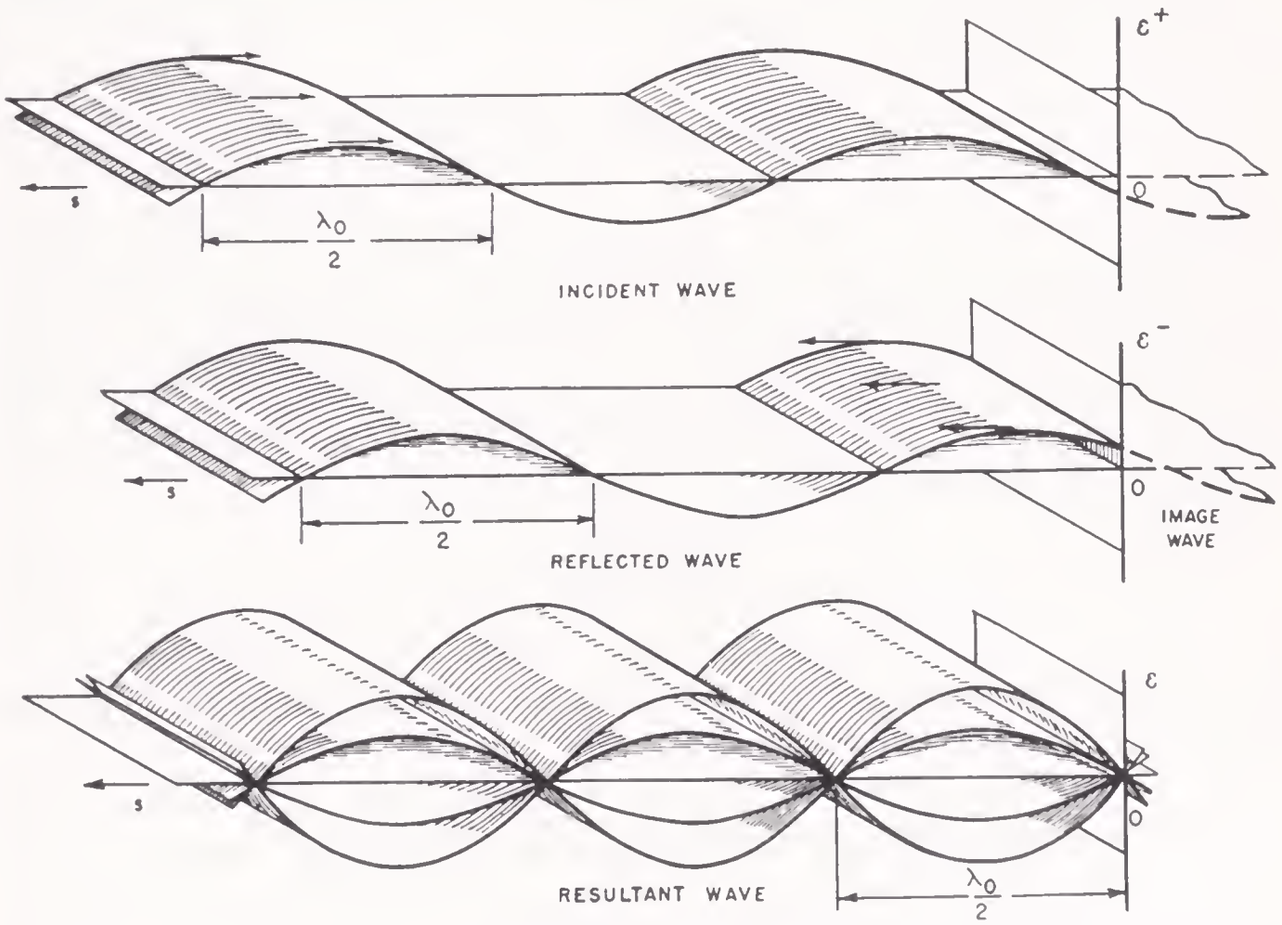
$$\mathcal{E}^+ = \mathcal{E}_m{}^+ \cos (\omega t + \beta_0 s) \tag{14}$$



(A) PARALLEL-STRIP LINE AND REFLECTOR



(B) STANDING WAVES OF VOLTAGE AND CURRENT



(C) ELECTRIC-FIELD WAVES

FIG. 5. Standing waves in parallel-strip line with reflector perpendicular to direction of propagation.

represents the incident wave of electric-field intensity, then

$$\mathcal{E}^- = -\mathcal{E}_m{}^+ \cos(\omega t - \beta_0 s) \tag{15}$$

must be the reflected wave in order that, at $s = 0$, $\mathcal{E} = \mathcal{E}^+ + \mathcal{E}^- = 0$. The resultant electric-field intensity is therefore the standing wave

$$\mathcal{E} = \mathcal{E}^+ + \mathcal{E}^- = \mathcal{E}_m{}^+[\cos(\omega t + \beta_0 s) - \cos(\omega t - \beta_0 s)]$$

or

$$\mathcal{E} = -(2\mathcal{E}_m{}^+ \sin \beta_0 s) \sin \omega t = 2(\mathcal{E}_m{}^+ \sin \beta_0 s) \cos(\omega t + 90°) \tag{16}$$

The quantity $2\mathcal{E}_m{}^+ \sin \beta_0 s$ is independent of time and specifies the peak value of the sine-wave variation of the total electric-field intensity at points $\beta_0 s$ electrical degrees from the reflector.

A convenient way of visualizing the electric-field waves is to consider the reflecting surface as a mirror and to regard the reflected wave as the image of the incident wave. Thus, as the incident wave moves toward the reflector, an image wave which appears to have its source behind the reflector moves away from the reflector. The incident and image waves combine in the region in front of the reflector to produce the resultant electric-field-intensity pattern. This concept of image waves could have been applied equally well to voltage and current waves in Chap. VII.

Accompanying the incident electric-field wave is an incident magnetic-field wave of intensity

$$\mathcal{H}^+ = \mathcal{H}_m{}^+ \cos(\omega t + \beta_0 s) = \frac{\mathcal{E}_m{}^+}{Z_i} \cos(\omega t + \beta_0 s) \tag{17}$$

The direction of propagation of the reflected magnetic-field wave $\mathcal{H}^-$ is the same as that of the $\mathcal{E}^-$ wave. To maintain the proper directional relationship between the electric- and magnetic-field vectors of the reflected wave, the incident magnetic-field wave must be reflected *without* phase reversal. Thus, when the reference directions are those of Fig. 5*A*, the image wave of magnetic-field intensity is

$$\mathcal{H}^- = \frac{\mathcal{E}_m{}^+}{Z_i} \cos(\omega t - \beta_0 s) \tag{18}$$

and the total magnetic-field intensity $\mathcal{H}$ at any point in front of the reflecting plane is the standing wave

$$\mathcal{H} = \mathcal{H}^+ + \mathcal{H}^- = \frac{\mathcal{E}_m{}^+}{Z_i}[\cos(\omega t + \beta_0 s) + \cos(\omega t - \beta_0 s)]$$

or

$$\mathcal{H} = \left(\frac{2\mathcal{E}_m{}^+}{Z_i} \cos \beta_0 s\right) \cos \omega t \tag{19}$$

Equation 19 indicates that the peak value of the sinusoidal variations of $\mathfrak{H}$ changes with distance from the reflector in the same manner as does the peak value of the current wave in Fig. 5*B*. Comparison of Eqs. 16 and 19 shows that the sinusoidal variations of $\mathcal{E}$ and $\mathfrak{H}$ are 90 deg apart in phase.

The total voltage and current e and i in the standing-wave pattern, unlike voltage and current of a traveling wave, are not related by a constant ratio R_c. Rather, the e and i variations are 90 deg out of phase, and the relationship must be expressed as the ratio of the complex rms values E and I, which is (see Art. 3, Chap. VII)

$$\frac{E}{I} = Z = jR_c \tan \beta s$$

Z being the input impedance of a section of line of length s terminated in a short circuit. In like manner, the variations of electric-field intensity and magnetic-field intensity of the waves in Fig. 5 differ 90 deg in phase, and the relationship of $\mathcal{E}$ to $\mathfrak{H}$ at any cross section of the line is expressed by the *wave impedance*

$$\mathcal{Z} = j\mathcal{Z}_i \tan \beta s \tag{20}$$

which is a complex number. The angle of $\mathcal{Z}$ is +90 deg (when tan βs is positive), indicating that, for the reference directions of Fig. 5, $\mathcal{E}$ leads $\mathfrak{H}$ by 90 deg, as is specified by Eqs. 16 and 19. The amplitude ratio is $\mathcal{Z}_i$ tan βs, again in agreement with Eqs. 16 and 19. Quantity $\mathcal{Z}$ is the input wave impedance for the length s of parallel-strip line terminated by the short-circuiting reflector.

Because the input wave impedance $\mathcal{Z}$ is merely a constant w/h times the input impedance Z, the Smith chart may be employed with $\mathcal{Z}$ in exactly the same manner as with Z, except that per-unit values are based upon $\mathcal{Z}_i$ instead of upon R_c. In this way, the input wave impedance at any cross section may be found for a line terminated in a short-circuiting conductor or in any other device of known wave impedance. Wave impedances may be thought of as being resistive, capacitive, or inductive, depending upon the associated phase angle between $\mathcal{E}$ and $\mathfrak{H}$. Similarly, a cross section in which $\mathcal{E}$ is zero but $\mathfrak{H}$ is not zero, corresponding to $\mathcal{Z} = 0$, may be thought of as a short circuit, whereas a zero value of $\mathfrak{H}$ and a nonzero value of $\mathcal{E}$ at a cross section corresponds to infinite $\mathcal{Z}$, or an open circuit. In Fig. 5, a short circuit in input impedance is indicated at the reflecting plane and at intervals of a half wavelength from the reflection; an open circuit corresponds to positions an odd number of quarter wavelengths from the reflector. The short-circuit condition of zero $\mathcal{E}$ can be realized with high accuracy by use of a conducting reflector of dimensions greater than the cross-sectional dimension of the parallel-strip line. The

open-circuit condition of zero $\mathcal{H}$ is only imperfectly produced by an open-ended line, because fringing fields produce the effect of a terminating capacitance. For a single frequency, the open-circuit condition is perhaps best obtained by means of a short circuit located a quarter wavelength beyond the desired open-circuit point.

Because the dielectric and conductors in Fig. 5 are assumed to be without loss, the energy consumed by the terminated line section is zero, and the average power input to the line must be zero. Calculations based upon the Poynting vector agree with this conclusion. Both field vectors of the standing-wave pattern lie in the transverse plane, and therefore the Poynting vector is perpendicular to the transverse plane and specifies by its magnitude the power crossing a unit area of the transverse plane. The magnitude of the Poynting vector is

$$\mathcal{S} = \mathcal{E}\mathcal{H} = -(2\mathcal{E}_m^+ \sin \beta_0 s)\left(2\frac{\mathcal{E}_m^+}{\mathcal{Z}_i}\cos \beta_0 s\right)\sin \omega t \cos \omega t$$

$$\mathcal{S} = \left[\frac{(\mathcal{E}_m^+)^2}{\mathcal{Z}_i}\sin 2\beta_0 s\right]\sin 2\omega t \tag{21}$$

Thus the power crossing any area of a transverse plane varies sinusoidally with time at a frequency twice the frequency of the fields. The direction of power flow is toward the reflector for one half of the sin $2\omega t$ cycle and away from the reflector for the other half. The average of sin $2\omega t$ over any integral number of cycles is zero, and thus the average power transmitted by any cross-sectional area is zero. For cross sections spaced at intervals of a quarter wavelength from the reflector, sin $2\beta_0 s$ is zero, and the power transmitted is zero at every instant.

5. Oblique Reflection of Plane Transverse Waves. If the terminating reflector for the line of Fig. 5*A* is placed at an oblique angle, field patterns entirely different from those of Fig. 5*C* are obtained. The new patterns are typical of field patterns in rectangular waveguides. Determination of these patterns requires the use of the law of reflection and of the boundary conditions at the surface of a conductor.

The law of reflection for a plane boundary surface may be stated as follows:

Each ray of an electromagnetic wave reflected from a plane boundary between homogeneous materials lies in the plane of an incident ray and a normal to the boundary, and the angle between reflected ray and normal equals the angle between incident ray and normal.

This law applies no matter what materials are separated by the boundary plane, provided the relation between incident and reflected wave, whatever it may be, is the same at each point of the boundary plane. Proof of the law depends upon this uniformity of reflection conditions.

Figure 6A depicts a plane wave incident upon a plane boundary, the angle between a ray of the wave R_1O and a normal to the boundary ON being α. Wavefronts spaced at intervals of a wavelength λ_0 are represented by the lines drawn perpendicular to R_1O. Traces of these wavefronts in the boundary plane appear as points in the top view and as lines in the front view, and they are spaced at intervals $\lambda_0/\sin\alpha$. As the

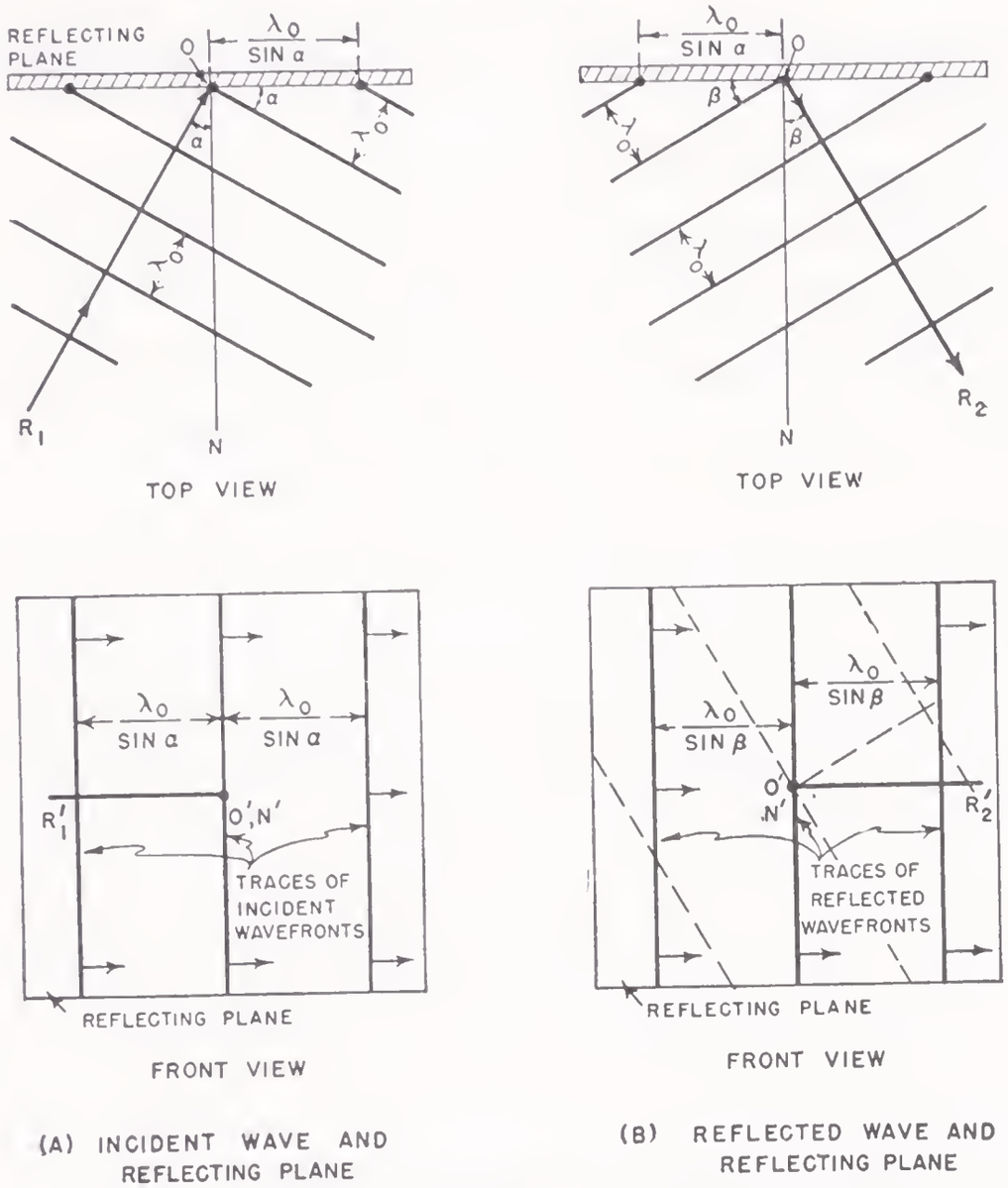


FIG. 6. Reflection of plane waves from plane boundary.

incident wave progresses toward the boundary, the traces in the boundary plane move from left to right.

The electric and magnetic fields of the incident wave vary from point to point in the boundary plane, but at any instant the fields are constant along each wavefront trace and execute a cycle of variation from one wavefront trace to the next. Because the conditions of reflection are the same at each point, the ratio of field intensities in the reflected wave to those in the incident wave must be the same at every point of the boundary plane. Thus the field intensities of the reflected wave must vary from point to point in the boundary plane in the same manner as those

of the incident wave. The wavefront traces of the reflected wave must coincide with the traces of Fig. 6*A* and must move in the same direction and at the same speed.

The dashed lines in Fig. 6*B* are drawn to show that unless a ray OR_2 of the reflected wave lies in the plane of R_1O and ON, coincidence of incident and reflected wavefront traces cannot be obtained. With OR_2 in the plane of ON and R_1O, the angles β and α must be equal in order that distances $\lambda_0/\sin\beta$ and $\lambda_0/\sin\alpha$ may be equal. Rays OR_2 and R_1O must lie on opposite sides of ON in order that wavefront traces for the two waves may move in the same direction.

The law of reflection specifies the direction of rays of the reflected wave but does not indicate the polarization of the wave nor the phase and magnitude of the fields of the wave. These conditions depend upon the nature of the materials separated by the boundary plane. If one of the materials is a perfect conductor—as assumed for the parallel strips and reflecting plate in Fig. 5*A*—the properties of the reflected wave may be determined from the following boundary conditions:

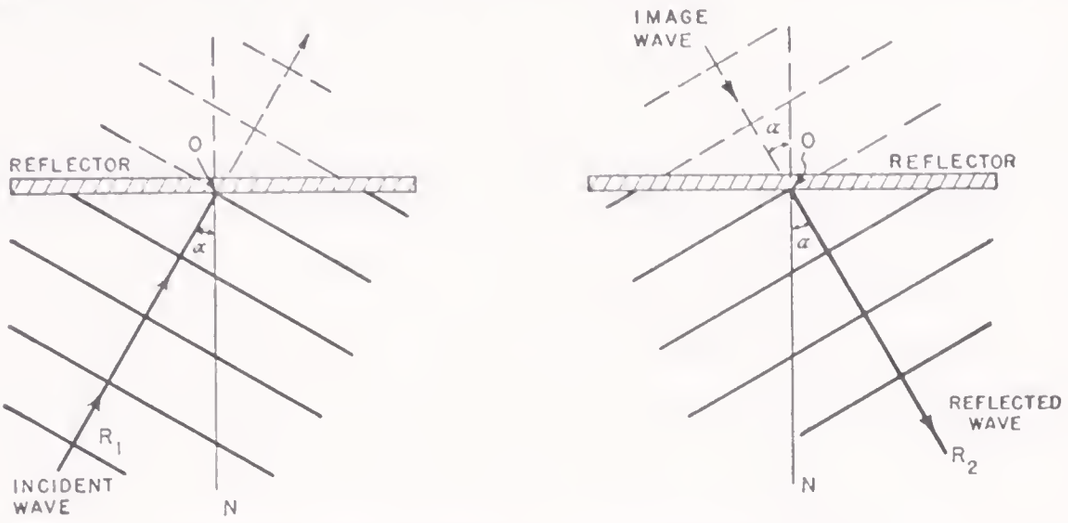
At the surface of a perfect conductor, there can be no component of electric field tangential to the surface and no component of a time-varying magnetic field perpendicular to the surface.

The electric-field condition has been used in Art. 4 as the field-quantity analogue of a circuit condition of zero voltage across a short circuit. This condition is necessary because a parallel electric field at the conductor surface would correspond to the impossible condition of a voltage difference between two points of a perfect conductor. The magnetic-field condition is the consequence of the electric-field condition and Faraday's law. If a time-varying magnetic field were to exist normal to the surface, the associated varying magnetic flux would cause an electromotive force around a closed path in the surface—a condition inconsistent with the requirement of zero tangential component of electric field at the surface.

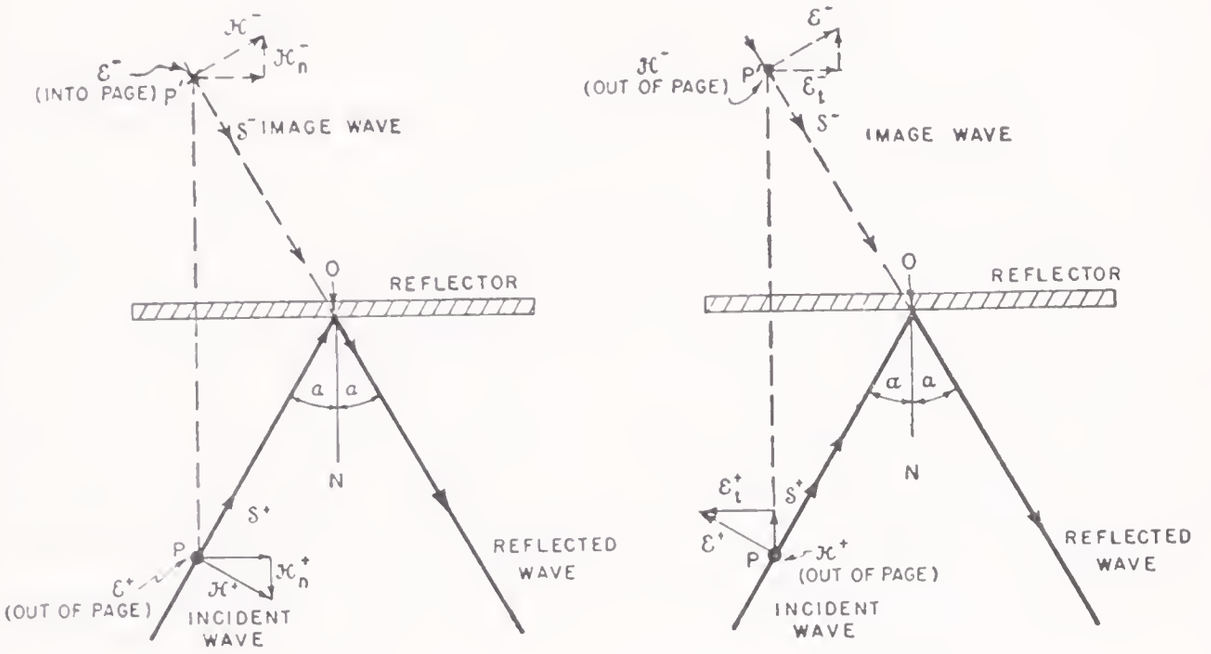
The boundary conditions do not preclude the existence of an electric field normal to a perfect conductor. Electric-field lines may terminate at charges on the surface of conductors, the charge per unit area at any point of the surface being equal to the density of electric-flux lines $\mathfrak{D} = \varepsilon\mathcal{E}$ terminating at the point. Similarly, a time-varying magnetic field parallel to a perfectly conducting surface is possible; and as explained in Art. 2, the parallel magnetic field is related to the current flowing along the surface of the conductor. The magnetic field is perpendicular to the direction of the current flow, and the field intensity $\mathfrak{IC}$ equals $\mathcal{J}$, the current per unit width of current path.

The law of reflection is justification for extending the concept of image waves to include reflection from a perfectly conducting surface placed at

an oblique angle to the rays of an incident wave. Figure 7A shows the relative positions of the incident, image, and reflected waves when a ray of the incident wave makes an angle α with a normal to the reflector. The image of the incident ray is located behind the reflector and makes an



(A) IMAGE WAVE DETERMINED BY LAW OF REFLECTION



(B) OUT-OF-PHASE REFLECTION OF WAVE WITH $\mathcal{E}$ FIELD PARALLEL TO REFLECTOR

(C) IN-PHASE REFLECTION OF WAVE WITH $\mathcal{H}$ FIELD PARALLEL TO REFLECTOR

FIG. 7. Image waves and the effect of polarization.

angle α with the image of the normal. Extension of the image ray to the incident-wave side of the reflector yields a ray of the reflected wave. Wavefronts of the waves are drawn perpendicular to the rays.

The image wave may be specified more definitely when the polarization of the incident wave is known. For example, in Fig. 7*B* the incident

wave is assumed to be linearly polarized with the electric field parallel to the reflector. The requirement of zero electric field tangential to the reflecting surface necessitates that the electric fields of the incident and image waves have equal intensities and opposite directions at point O and hence at any pair of points, such as P and P', that are images of each other. Absence of an electric field at the boundary ensures that the incident magnetic field is reflected in a manner that causes the normal components of the incident and reflected magnetic fields to add to zero at the boundary. As the result of satisfying these conditions, the incident wave is reflected without change of polarization, the incident electric field and the normal component of the incident magnetic field are reflected without change in amplitude but with a reversal of phase, and the tangential magnetic-field component is reflected without change of amplitude or phase. Relative directions and amplitudes of the fields of the incident and image waves are indicated at points P and P' in Fig. 7*B*.

When the *magnetic* field of the incident wave is parallel to the reflector, as in Fig. 7*C*, the same boundary conditions must be satisfied as before. The vector diagrams at points P and P' in Fig. 7*C* show that a change in direction of polarization (direction of the $\mathcal{E}$ vector) occurs on reflection, that the incident magnetic field and the normal component of the incident electric field are reflected without change in amplitude or phase, and that the tangential component of the incident electric field is reflected without change in amplitude but with a reversal of phase. Transverse waves having other polarizations may be divided into two component waves, to one of which Fig. 7*B* applies and to the other of which Fig. 7*C* applies.

6. Field Patterns Obtained by Oblique Reflection. The field patterns in front of a plane conductor placed at an oblique angle to the direction of wave propagation in a parallel-strip line may be determined through use of the principles developed in Art. 5. The conducting plate in Fig. 8*A* is perpendicular to the parallel strips but makes an angle α with the incident wavefronts. In accordance with the reflection law, rays of incident and reflected waves must make equal angles with the normal to the reflector, as shown in Fig. 8*A*. The reflected wave is therefore propagated along a line placed at an angle 2α with respect to the path of the incident wave.

The resultant field patterns may be conveniently described in terms of the rectangular-coordinate system shown in Fig. 8*A*. The x, z coordinate plane of this system is drawn in Fig. 8*B*, and the wavefronts of the incident wave are shown advancing toward the reflecting plane. Wavefronts of the image wave are represented in Fig. 8*C*. Note that the angle between a ray and a normal, or between a wavefront and the reflector, is α for both the incident and image waves.

If s_1 is a distance measured to the left from the origin along a ray of the

incident wave, the electric-field intensity of the incident wave is

$$\mathcal{E}^+ = \mathcal{E}_m^+ \cos(\omega t + \beta_0 s_1) \tag{22}$$

Similarly, if s_2 represents a distance along the ray of the image wave measured diagonally upward to the left, the field intensity $\mathcal{E}^-$ of the image wave is

$$\mathcal{E}^- = \mathcal{E}_m^+ \cos(\omega t - \beta_0 s_2) \tag{23}$$

These equations apply for the reference directions of $\mathcal{E}^+$ and $\mathcal{E}^-$ indicated at points P and P' in Figs. 8*B* and *C*—$\mathcal{E}^+$ is positive when directed

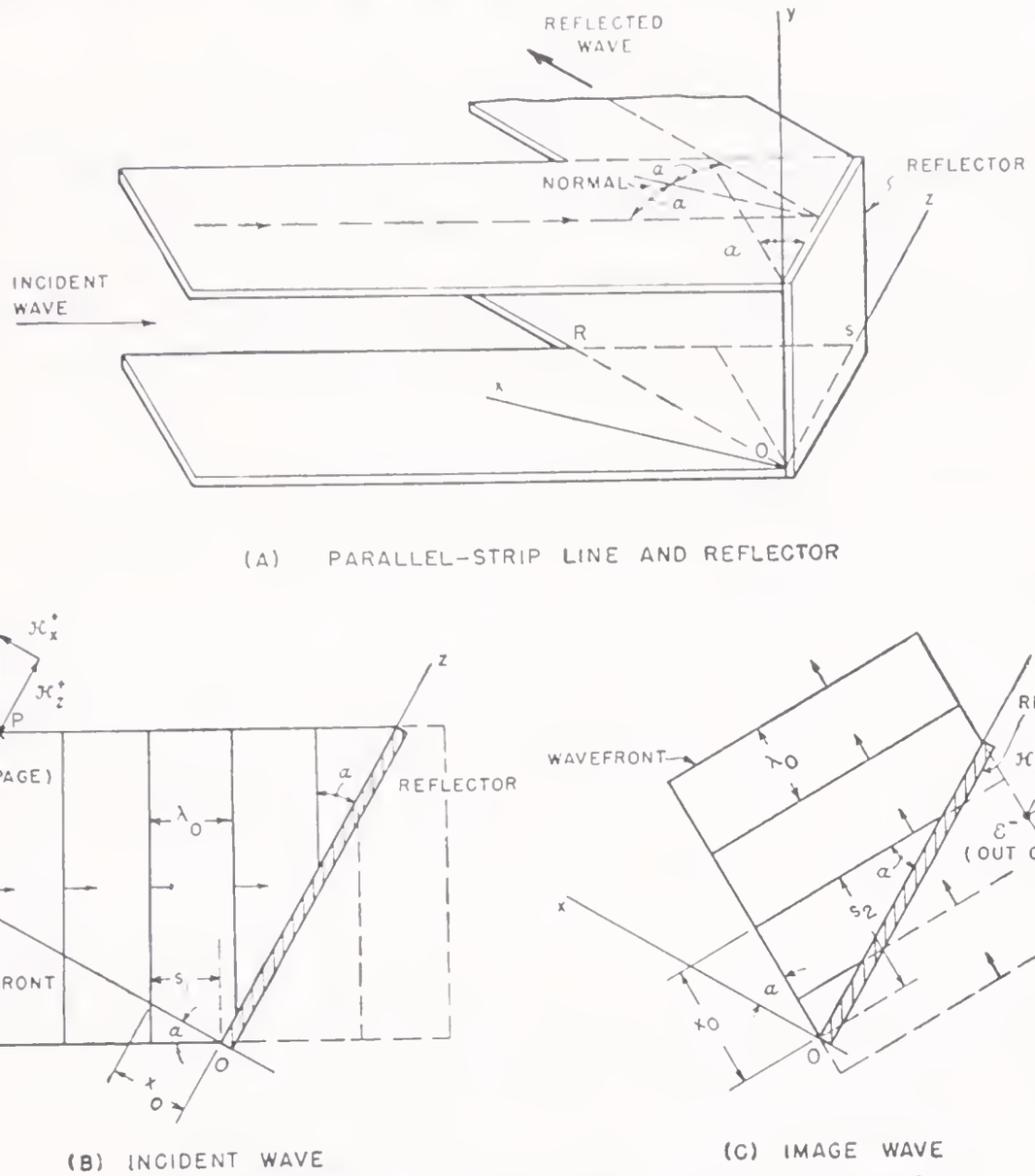


FIG. 8. Reflection from plane conductor at oblique angle.

downward, and $\mathcal{E}^-$ is positive when directed upward. At the origin ($s_1 = s_2 = 0$), $\mathcal{E}^- = \mathcal{E}^+$ and boundary conditions are therefore satisfied at this point.

In order to combine the incident and reflected electric-field intensity waves, s_1 and s_2 must be expressed in terms of the coordinates x and z. Distance s_1 is the normal distance from the origin to a particular incident wavefront. The trace of the plane of this wavefront in the x, z plane is

the line

$$x = x_0 + z \tan \alpha = \frac{s_1}{\cos \alpha} + z \frac{\sin \alpha}{\cos \alpha}$$

Thus $s_1 = x \cos \alpha - z \sin \alpha$. Similarly, it may be shown that distance $s_2 = x \cos \alpha + z \sin \alpha$. Substitution of these expressions for s_1 and s_2 in Eqs. 22 and 23 yields the value of the incident and reflected electric fields at any point in the x, z plane; that is,

$$\mathcal{E}^+ = \mathcal{E}_m{}^+ \cos [\omega t + (\beta_0 \cos \alpha)x - (\beta_0 \sin \alpha)z] \tag{24}$$

and

$$\mathcal{E}^- = \mathcal{E}_m{}^+ \cos [\omega t - (\beta_0 \cos \alpha)x - (\beta_0 \sin \alpha)z] \tag{25}$$

Note that at the surface of the reflector ($x = 0$), the total electric field, which is the difference between $\mathcal{E}^+$ and $\mathcal{E}^-$, is zero for all values of z and t. Boundary conditions are therefore satisfied over the entire reflecting surface.

In the space above triangle ORS in Fig. 8*A*, both incident and reflected waves are present, and the resultant electric-field intensity $\mathcal{E}$ in this region is $\mathcal{E}^+ - \mathcal{E}^-$; thus

$$\mathcal{E} = \mathcal{E}_m{}^+ \cos [\omega t - (\beta_0 \sin \alpha)z + (\beta_0 \cos \alpha)x] - \mathcal{E}_m{}^+ \cos [\omega t - (\beta_0 \sin \alpha)z - (\beta_0 \cos \alpha)x] \tag{26}$$

Through use of the identity

$$\cos (\phi + \theta) - \cos (\phi - \theta) = -2 \sin \phi \sin \theta$$

where $\phi = [\omega t - (\beta_0 \sin \alpha)z]$ and $\theta = (\beta_0 \cos \alpha)x$, Eq. 26 becomes

$$\mathcal{E} = \{-2\mathcal{E}_m{}^+ \sin [(\beta_0 \cos \alpha)x]\} \sin [\omega t - (\beta_0 \sin \alpha)z]$$

or in terms of a cosine function

$$\mathcal{E} = \{2\mathcal{E}_m{}^+ \sin [(\beta_0 \cos \alpha)x]\} \cos [\omega t + 90° - (\beta_0 \sin \alpha)z] \tag{27}$$

Equation 27 represents a wave of peak value $2\mathcal{E}_m{}^+ \sin [(\beta_0 \cos \alpha)x]$ traveling parallel to the z axis and, hence, parallel to the reflecting surface. The peak value of the traveling wave is a function of x, the normal distance from the reflector, and varies between zero and $2\mathcal{E}_m{}^+$ in accordance with variations of $\sin [(\beta_0 \cos \alpha)x]$. Thus a *standing wave* is produced normal to the reflector, and as time goes on, this standing wave moves parallel to the reflecting surface.

Equation 27 indicates that the phase constant of the resultant electric-field wave traveling parallel to the reflector is different from the phase constant of the standing-wave pattern normal to the reflector. Both phase constants are also different from β_0, the phase constant of the incident and reflected waves. If β represents the phase constant of the

resultant wave in the z direction,

$$\beta = \beta_0 \sin \alpha \tag{28}$$

Similarly, if β_x is the phase constant of the standing wave in the x direction,

$$\beta_x = \beta_0 \cos \alpha \tag{29}$$

Thus in terms of β and β_x, the resultant electric-field intensity in the region near the reflector is

$$\mathcal{E} = (2\mathcal{E}_m{}^+ \sin \beta_x x) \cos (\omega t + 90^\circ - \beta z) \tag{30}$$

The curved-surface method of representing electric-field-intensity waves may be used to describe the electric-field pattern represented by Eq. 30. The incident and reflected wave components are shown in Figs. 9*A* and *B*, and the resultant pattern is given in Fig. 9*C*. Note that the field-intensity variations of each component wave in the reflecting plane are sinusoidal; but because the two intensities are equal in amplitude and opposite in direction, the resultant electric field at the reflector is zero. The component waves travel in the directions indicated by the arrows, and the resultant field pattern travels parallel to the z axis.

Since wavelength is 2π divided by the phase constant, the wavelengths in the z and x directions in Fig. 9*C* are different from each other and from λ_0. In the z direction, the wavelength λ is

$$\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{\beta_0 \sin \alpha} = \frac{\lambda_0}{\sin \alpha} \tag{31}$$

and in the x direction, λ_x is

$$\lambda_x = \frac{2\pi}{\beta_x} = \frac{\lambda_0}{\cos \alpha} \tag{32}$$

These wavelengths are marked in Fig. 9*C*.

A physical reason for the nature of the field pattern illustrated in Fig. 9*C* and for the wavelength relationships of Eqs. 31 and 32 may be seen by again considering Figs. 8*B* and *C*. To an observer capable of exploring the electric-field variations along a single line parallel to the reflector, each component wave would appear as a wave traveling in the z direction. The distance between successive crests of either wave in the z direction is $\lambda_0/\sin \alpha = \lambda$. The combination of these two component waves is a wave of length λ moving in the same direction. The amplitude of the resultant wave depends upon the distance of line of observation from the reflector. To an observer confined to a line perpendicular to the reflector, the incident wave would seem to move toward the reflector and the reflected wave away from it. The wavelengths of both waves would appear to be

$\lambda_0/\cos\alpha = \lambda_x$, as indicated in Fig. 9. In exploring along the perpendicular line, therefore, the observer would encounter waves similar to those of Fig. 5 except for the longer wavelength and would regard the resultant field pattern as a standing wave with zeros at the reflector and at $\lambda_x/2$ intervals from the reflector.

The effect on λ and λ_x of varying the angle α may be noted from Eqs. 31 and 32. As α approaches zero, the wavelength λ measured parallel to

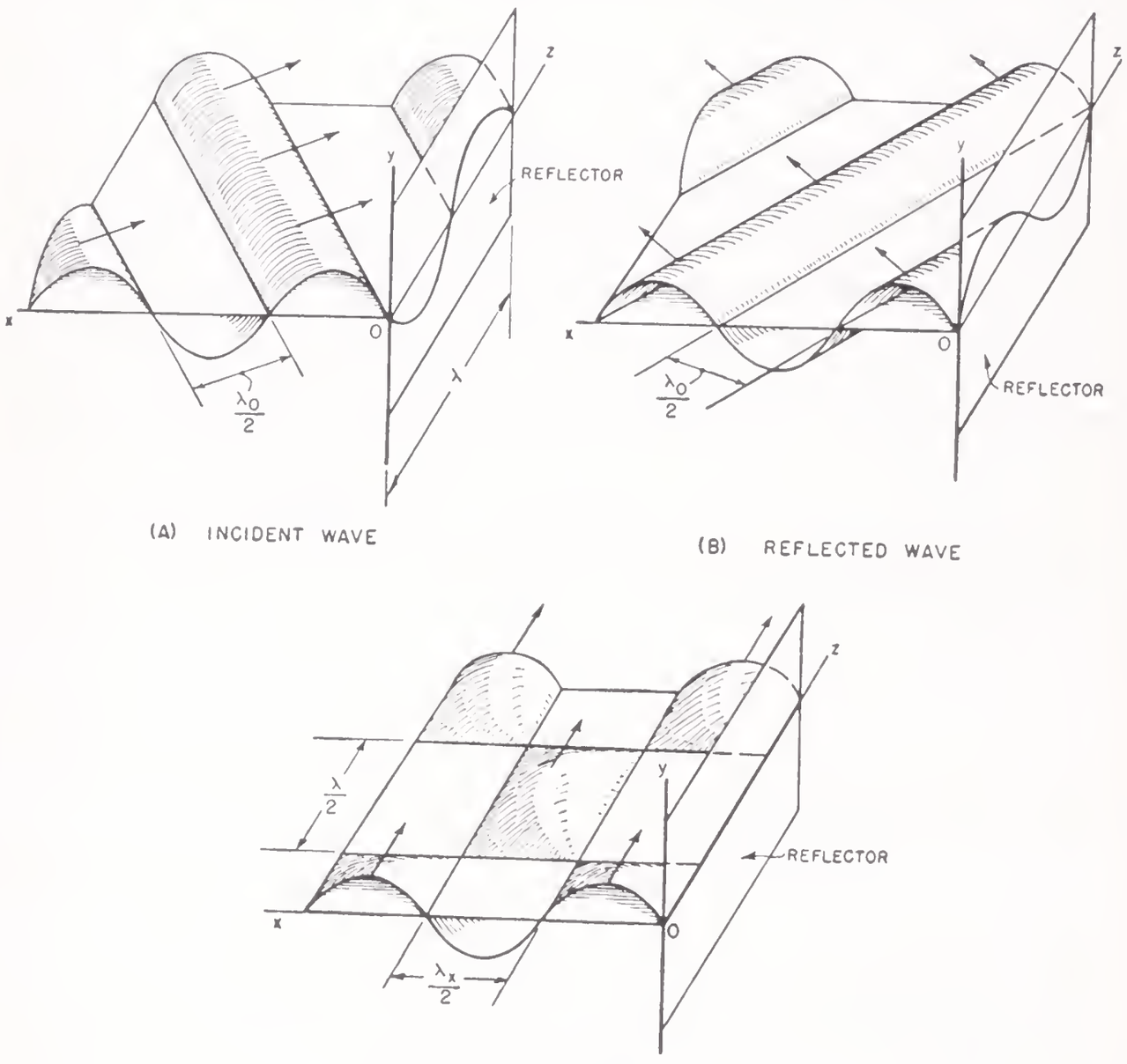


FIG. 9. Representation of electric-field pattern near plane reflector.

the reflector increases without limit, while the wavelength λ_x in the normal direction approaches λ_0. Thus the condition of normal incidence illustrated in Fig. 5 is approached. On the other hand, as α approaches 90 deg (grazing incidence), the length of the standing wave in the normal direction increases without limit, while the length of the waves traveling parallel to the z axis approaches λ_0. The single-wave field patterns of Fig. 4 are therefore approached as a limit. Combination of Eqs. 28 and 29 shows that for every value of α the waves in the normal and parallel

directions are related so that $\beta^2 + \beta_x^2 = \beta_0^2$ and hence so that

$$\frac{1}{\lambda^2} + \frac{1}{\lambda_x^2} = \frac{1}{\lambda_0^2} \tag{33}$$

Associated with the incident and reflected waves of electric field in Figs. 8 and 9 are magnetic fields of intensities $\mathcal{H}^+ = \mathcal{E}^+/Z_i$ and $\mathcal{H}^- = \mathcal{E}^-/Z_i$. Dividing Eq. 24 by Z_i thus yields as the expression for $\mathcal{H}^+$

$$\mathcal{H}^+ = \frac{\mathcal{E}_m^+}{Z_i} \cos [\omega t + (\beta_0 \cos \alpha)x - (\beta_0 \sin \alpha)z]$$

or

$$\mathcal{H}^+ = \frac{\mathcal{E}_m^+}{Z_i} \cos (\omega t + \beta_x x - \beta z) \tag{34}$$

and dividing Eq. 25 by Z_i gives as the expression for $\mathcal{H}^-$

$$\mathcal{H}^- = \frac{\mathcal{E}_m^+}{Z_i} \cos (\omega t - \beta_x x - \beta z) \tag{35}$$

Reference directions for $\mathcal{H}^+$ and $\mathcal{H}^-$ are indicated at points P and P' in Figs. 8*B* and *C*.

Because the incident and reflected magnetic fields are in different directions, the resultant magnetic-field intensity cannot be determined by a direct addition of Eqs. 34 and 35. Instead, the incident and reflected magnetic-field intensities are first resolved into components parallel and perpendicular to the reflector. As indicated at points P and P' in Figs. 8*B* and *C*, the components of $\mathcal{H}^+$ and $\mathcal{H}^-$ in the z direction are $\mathcal{H}_z^+ = \mathcal{H}^+ \cos \alpha$ and $\mathcal{H}_z^- = \mathcal{H}^- \cos \alpha$, and total magnetic intensity in the z direction is $\mathcal{H}_z$.

$$\begin{aligned} \mathcal{H}_z &= \mathcal{H}^+ \cos \alpha + \mathcal{H}^- \cos \alpha \\ &= \frac{\mathcal{E}_m^+}{Z_i} \cos \alpha [\cos (\omega t + \beta_x x - \beta_z) + \cos (\omega t - \beta_x x - \beta z)] \end{aligned} \tag{36}$$

Through use of the trigonometric identity

$$\cos (\phi + \theta) + \cos (\phi - \theta) = 2 \cos \phi \cos \theta$$

Eq. 36 simplifies to

$$\mathcal{H}_z = \left(2 \frac{\mathcal{E}_m^+}{Z_i} \cos \alpha \cos \beta_x x\right) \cos (\omega t - \beta z) \tag{37}$$

In the x direction, the component magnetic-field intensities are

$$\mathcal{H}_x^+ = \mathcal{H}^+ \sin \alpha$$

and $\mathcal{H}_x^- = \mathcal{H}^- \sin \alpha$, and the resultant intensity is $\mathcal{H}_x = \mathcal{H}_x^+ - \mathcal{H}_x^-$.

Thus

$$\begin{aligned}\mathcal{H}_x &= \mathcal{H}^+ \sin \alpha - \mathcal{H}^- \sin \alpha \\ &= \frac{\mathcal{E}_m^+ \sin \alpha}{Z_i} [\cos (\omega t + \beta_x x - \beta z) - \cos (\omega t - \beta_x x - \beta z)]\end{aligned}$$

Upon simplification, this expression becomes

$$\mathcal{H}_x = \left(2 \frac{\mathcal{E}_m^+}{Z_i} \sin \alpha \sin \beta_x x\right) \cos (\omega t + 90° - \beta z) \tag{38}$$

Equations 37 and 38 indicate that the magnetic field, like the electric field, has the appearance of a standing wave along a line perpendicular to the reflector and of a traveling wave along a line parallel to the reflector. The normal component of the magnetic field $\mathcal{H}_x$ is zero at the reflector, as may be seen by setting $x = 0$ in Eq. 38, and zero and maximum values of $\mathcal{H}_x$ occur alternately at regular intervals along normals to the reflector where $\sin \beta_x x$ is zero and maximum. Since the amplitude of $\mathcal{H}_x$ and $\mathcal{E}$ depends upon the factor $\sin \beta_x x$, the electric field and normal component of the magnetic field are zero at the same distances from the reflector. On the other hand, comparison of Eqs. 30 and 37 shows that $\mathcal{H}_z$ is maximum where the electric field is zero. This condition results because the magnitude of $\mathcal{H}_z$ depends upon the factor $\cos \beta_x x$ whereas the amplitude of $\mathcal{E}$ depends upon $\sin \beta_x x$.

Because the resultant magnetic field lies outside planes parallel to the reflector, the field may be represented more conveniently by the straight-line method employed in Fig. 4*B* than by diagrams similar to those of Fig. 9. Thus the lines in Fig. 10*A* represent variations in the incident magnetic field at one instant in the region near the reflector. Because this figure is a top view of a section between conductors, the electric-field lines appear only as dots and crosses. The field pattern of the reflected wave is shown in Fig. 10*B*, and the resultant field is illustrated in Fig. 10*C*. Observe in Fig. 10*C* that the resultant electric field is *transverse*, since it is confined wholly to planes which are at right angles to the direction of propagation. The magnetic field, on the other hand, is not transverse because of its component in the z direction. The resultant wave is therefore a *plane transverse-electric wave*. Waves of this type are important not only because they may be formed when plane transverse electromagnetic waves impinge upon plane reflectors but also because they may be propagated through rectangular waveguides.

7. The $TE_{1,0}$ Wave in a Rectangular Waveguide. Examination of Fig. 10*C* shows that, when plane transverse waves are reflected from a conductor oblique to the rays of the wave, the resultant electric-field intensity is zero not only at the surface of the conducting plane but also in

planes parallel to the conductor spaced at intervals of $\lambda_x/2$. The location of the electric field may be seen from Fig. 9*C* or 10*C*. It would seem possible to insert a conducting surface at any nodal plane without change of the field patterns, because boundary conditions would still be satisfied.

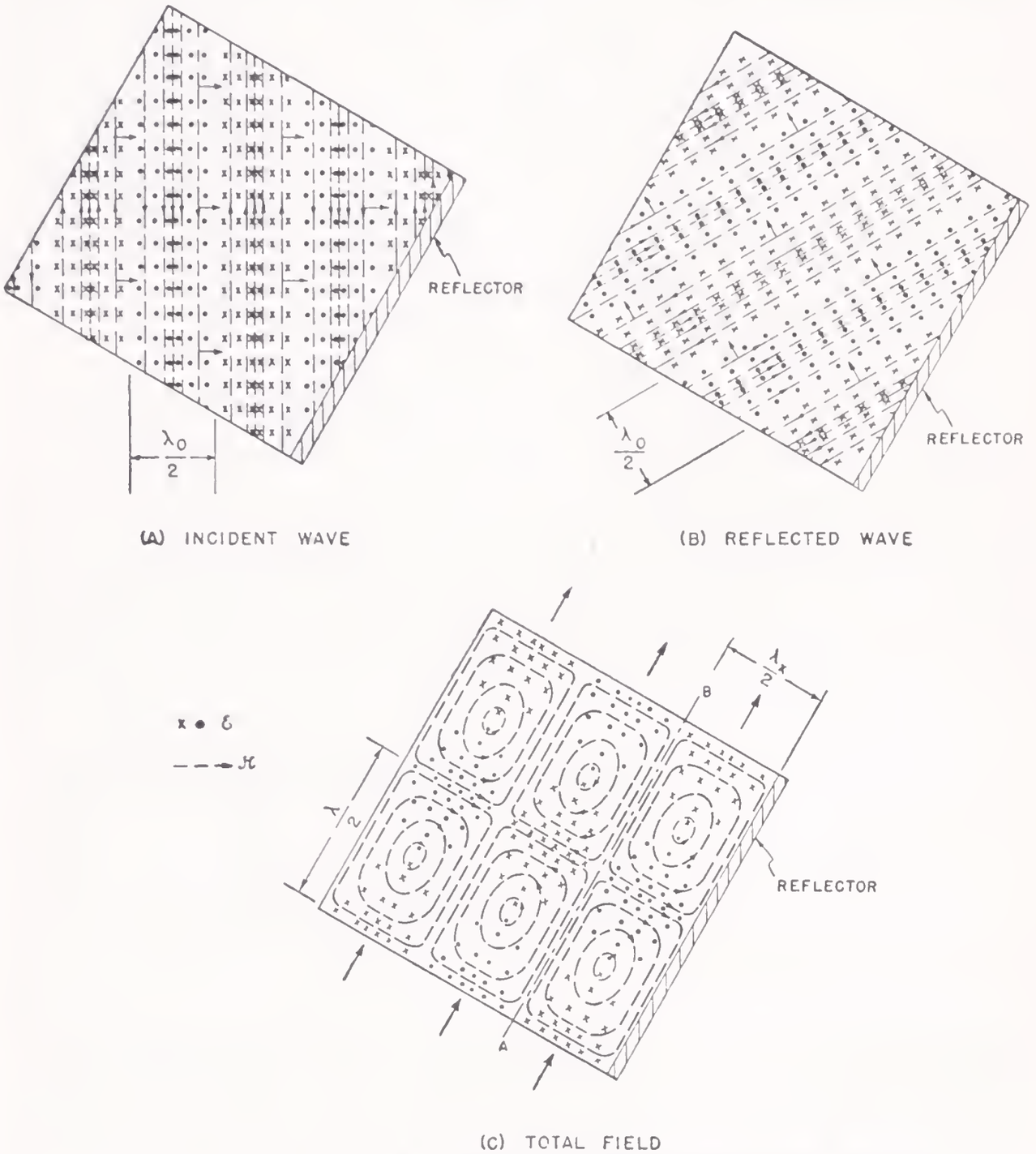
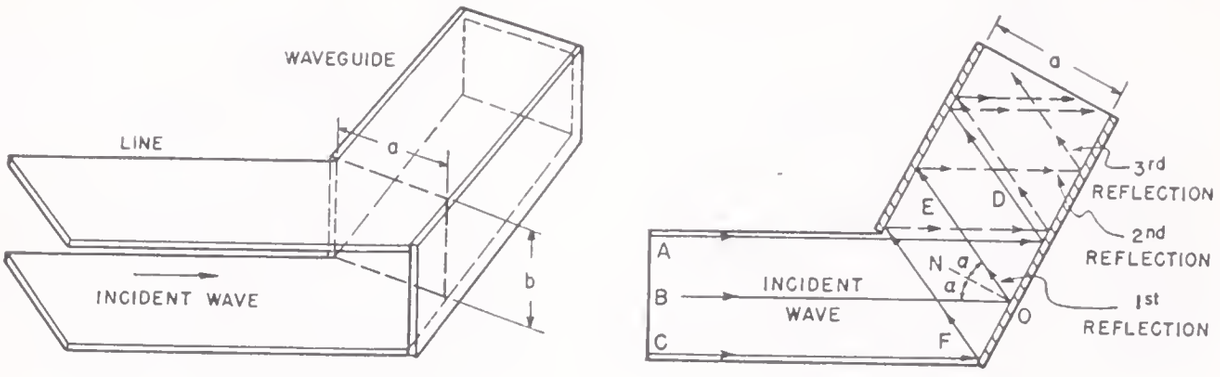


FIG. 10. Representation of electric and magnetic fields near reflector.

Addition of such a conductor to the parallel-strip line and reflector of Fig. 8*A* forms the rectangular waveguide of Fig. 11*A*.

The geometry of the rays of the component waves within the guide is shown by the drawing of Fig. 11*B*. Rays of the incident wave delivered by the parallel-strip line are indicated by solid lines *A*, *B*, and *C*, and the corresponding rays of the reflected wave by lines *D*, *E*, and *F*. Incident and reflected waves are at an angle α with *ON*, the normal to the reflector.

Rays D, E, and F undergo a second reflection at the surface of the left-hand reflecting conductor and thus produce the second-reflection rays shown by dashed lines. Third, fourth, and higher order reflections occur until the entire length of the waveguide is filled with two sets of plane transverse waves which cross the guide in opposite directions. Each set



(A) PARALLEL-STRIP LINE AND WAVEGUIDE (B) HORIZONTAL SECTION SHOWING RAYS OF INCIDENT AND REFLECTED WAVES

FIG. 11. Formation of rectangular waveguide from two-strip line with reflectors.

of waves travels at an angle α with the normal to the guide, and each produces the other wave by reflection at a side wall.

Because the two sets of waves are identical with the incident and reflected waves of Figs. 10A and B, the resultant wave has the pattern shown in Fig. 10C. In order that boundary conditions may be satisfied,

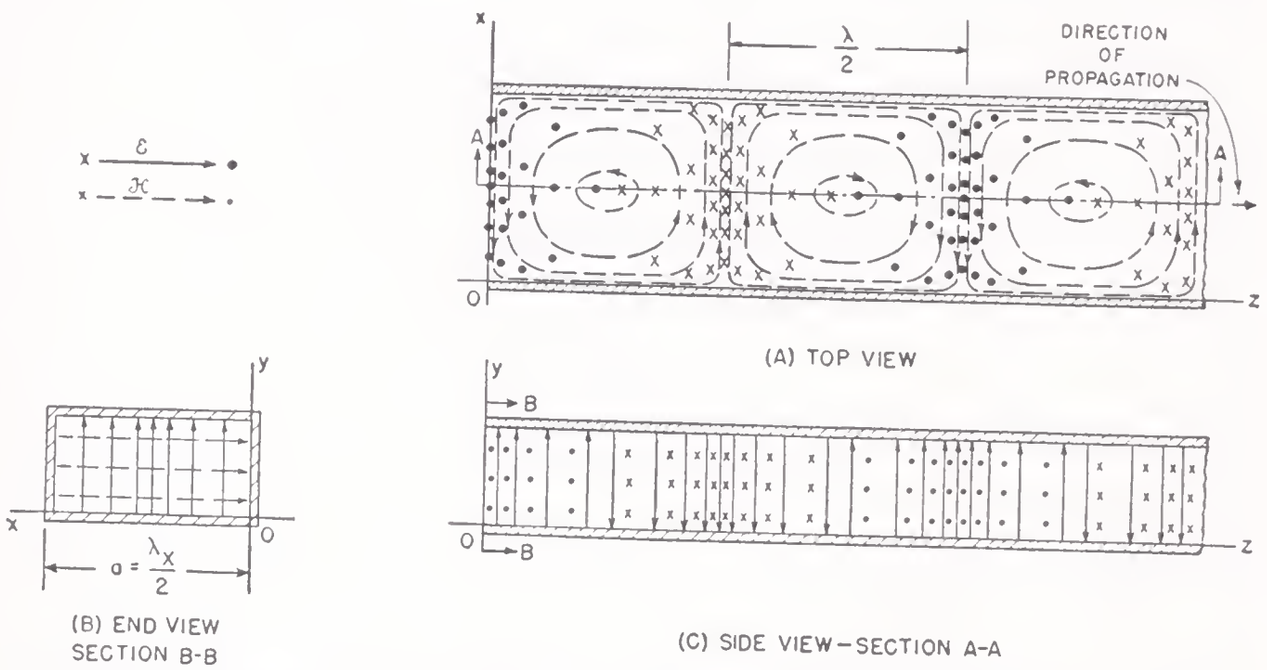


FIG. 12. Electric and magnetic field of the $TE_{1,0}$ wave in rectangular waveguide.

the guide dimension a must equal a multiple of $\lambda_x/2$. When $a = \lambda_x/2$, the fields in the waveguide are represented by the portion of Fig. 10C between line AB and the reflector. This portion of the pattern is shown in relation to the waveguide in Fig. 12.

As was pointed out in Art. 6, the pattern may be considered as a stand-

ing wave in the x direction, which is across the guide, and a traveling wave in the z direction, along the guide. Because a wave of this type in a guide is of great importance, it is convenient to think of it not merely as the combination of two plane transverse waves, but as a new wave type. It is called the $TE_{1,0}$ wave of a rectangular guide. The letters TE indicate that the electric field, but not the magnetic field, is transverse to the direction of propagation (see Fig. 12A). The subscripts 1 and 0 indicate the number of half-cycle variations of the electric-field intensity along the a and b dimensions of the guide, respectively, in accordance with a con-

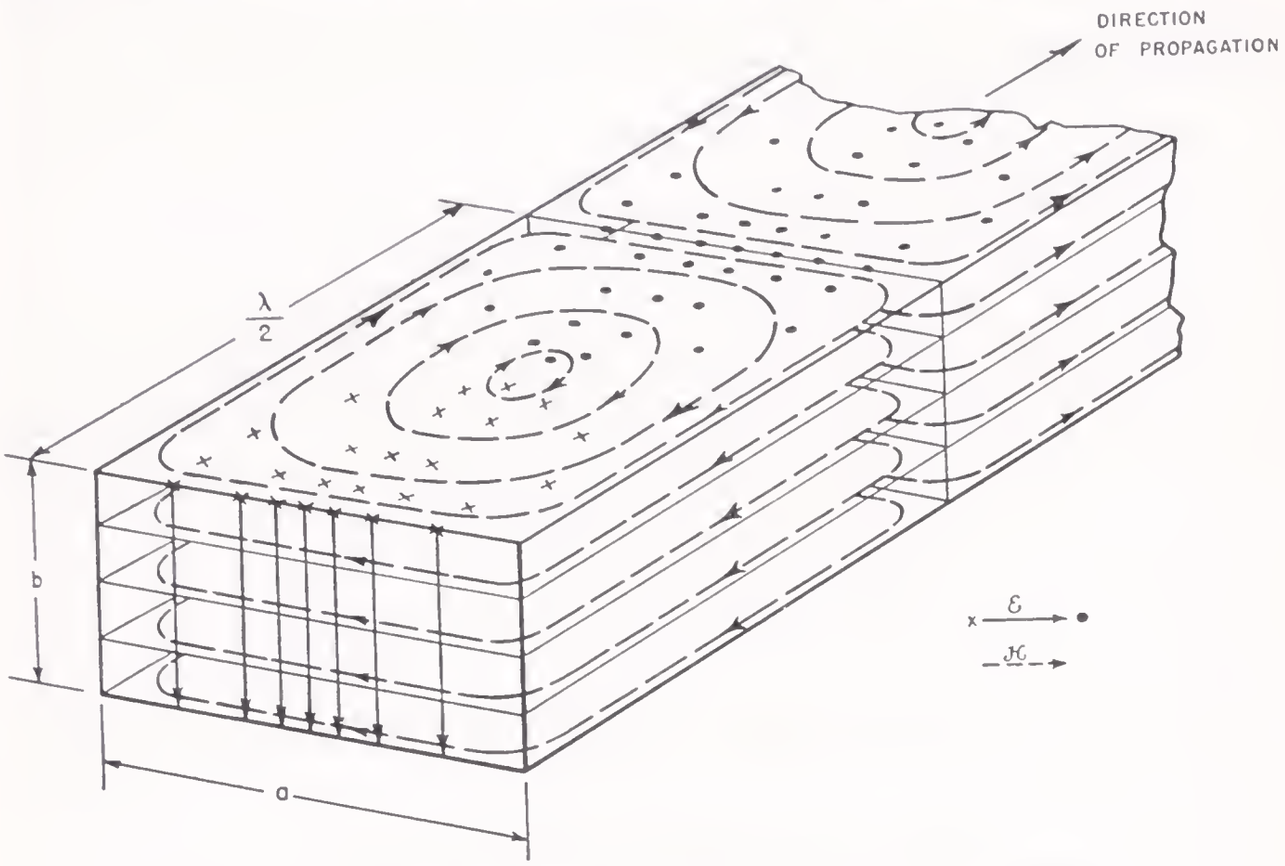


FIG. 13. Pictorial representation of field pattern of $TE_{1,0}$ wave.

vention explained in Art. 10. A pictorial representation of the $TE_{1,0}$ field pattern is shown in Fig. 13.

The requirement that a must equal $\lambda_x/2$ to obtain a $TE_{1,0}$ wave imposes a relation among the width a of the waveguide, the wavelength λ_0 of the component transverse wave, and the angle α at which the transverse waves enter the guide. According to Eq. 32, p. 567, the x-direction wavelength λ_x is $\lambda_0/\cos\alpha$. Since λ_x must equal $2a$,

$$\cos\alpha = \frac{\lambda_0}{2a} \tag{39}$$

Thus the frequency of the wave source (which fixes λ_0) and the guide width together determine the angle α and the geometry of the fields in the guide. As the source frequency is increased, λ_0 and $\cos\alpha$ decrease and α approaches 90 deg; that is, the higher the frequency of the source, the

more nearly the component transverse waves travel along the axis of the guide. As the source frequency is decreased, λ_0 increases, and the component waves travel more nearly perpendicular to the axis of the guide. When the value $\lambda_0 = 2a$ is reached, $\cos \alpha = 1$, $\alpha = 0$, and the pattern in the waveguide becomes merely a standing wave between the side walls of the guide. For longer wavelengths, Eq. 39 cannot be satisfied. Thus the waveguide is a form of high-pass filter capable of transmitting only those waves having a length less than the cutoff wavelength

$$\lambda_{co} = 2a \tag{40}$$

and a frequency greater than the cutoff frequency $f_{co} = c/\lambda_{co}$.

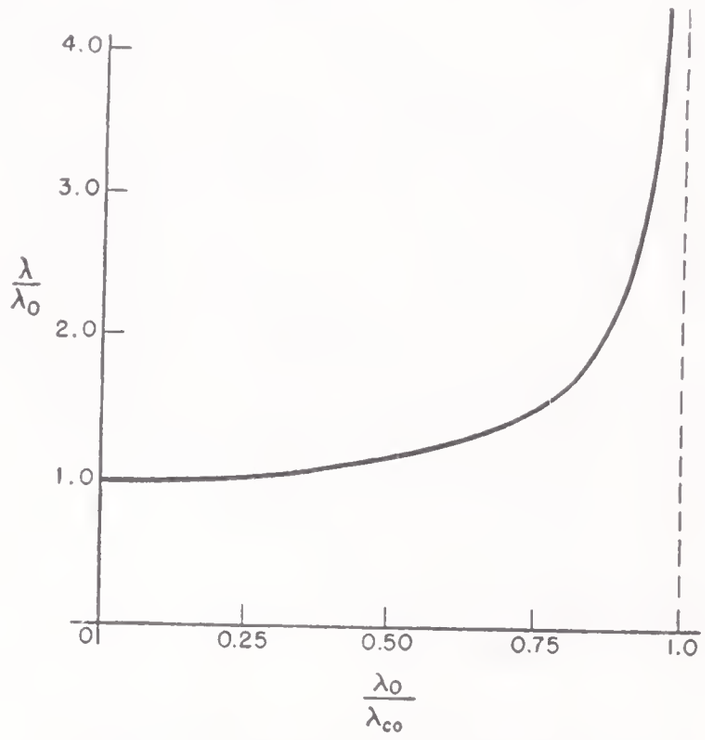


FIG. 14. Variation of λ/λ_0 with λ_0/λ_{co}.

The length of the $TE_{1,0}$ wave along the axis of the guide, designated by λ, is called the *guide wavelength*, and according to Eq. 31, p. 567, it is equal to $\lambda_0/\sin \alpha$. Consequently, the ratio of λ to λ_0 depends upon α, which in turn is a function of the relation of λ_0 to the cutoff wavelength, as follows:

$$\frac{\lambda}{\lambda_0} = \frac{1}{\sin \alpha} = \frac{1}{\sqrt{1 - \cos^2 \alpha}} = \frac{1}{\sqrt{1 - (\lambda_0/2a)^2}}$$

or

$$\frac{\lambda}{\lambda_0} = \frac{1}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} \tag{41}$$

A plot of Eq. 41 is drawn in Fig. 14. Observe that at very high frequencies (for which λ_0/λ_{co} approaches zero), the guide wavelength nearly equals the length of the component transverse waves λ_0. As the frequency is decreased, the guide wavelength increases (1) because λ_0

increases and (2) because the ratio λ/λ_0 increases. The change of λ/λ_0 is not very important if λ_0/λ_{co} is less than 0.5, but for values of λ_0 approaching λ_{co}, λ/λ_0 and the guide wavelength increase rapidly and approach infinity. Another form of Eq. 41, which can be obtained directly from Eq. 33, p. 569, and which is helpful in visualizing the wavelength relations, is

$$\frac{1}{\lambda^2} + \frac{1}{\lambda_{co}^2} = \frac{1}{\lambda_0^2} \tag{42}$$

The equations for the electric- and magnetic-field intensities developed in Art. 6 may be applied to describe the fields in the rectangular waveguide. Equation 30, p. 567, becomes, after β_x is replaced by $2\pi/\lambda_x = \pi/a$,

$$\mathcal{E} = \left(\mathcal{E}_m \sin \frac{\pi x}{a}\right) \cos (\omega t + 90° - \beta z) \tag{43}$$

In this equation, $\mathcal{E}_m = 2\mathcal{E}_m^+$, $\mathcal{E}_m^+$ being the peak value of electric-field intensity in either of the component transverse waves and $\mathcal{E}_m$ being the peak value of the electric-field intensity of the resultant $TE_{1,0}$ wave measured at the center of the guide. Since $\sin \alpha = \lambda_0/\lambda$, the transverse or x component of the magnetic field is (see Eq. 38, p. 570)

$$\mathcal{H}_x = \left(\frac{\mathcal{E}_m \lambda_0}{Z_i \lambda} \sin \frac{\pi x}{a}\right) \cos (\omega t + 90° - \beta z) \tag{44}$$

and the longitudinal or z component of the magnetic field (from Eq. 37, p. 569) is

$$\mathcal{H}_z = \left(\frac{\mathcal{E}_m \lambda_0}{Z_i \lambda_{co}} \cos \frac{\pi x}{a}\right) \cos (\omega t - \beta z) \tag{45}$$

The magnetic field at the surfaces of the conducting walls of the waveguide is parallel to these surfaces and, as explained in Art. 2, is related in a very definite way to currents in the waveguide walls. Figure 15 represents the magnetic field in the same way as did Fig. 13 and shows also the pattern of current flow in the guide walls. The walls are assumed to be perfect conductors, and therefore the current is confined to sheets of zero thickness at the inside surface of the walls. The linear density (current per unit path width) of current in these sheets is indicated by the density of current lines in Fig. 15, and the directions of current flow are indicated by the directions of the lines. Note that, as explained in Art. 2, the current lines are everywhere perpendicular to the magnetic-field lines at the conductor surface. The magnetic-field pattern may be thought of as being similar to the field in a series of approximately toroidal cores placed beside one another along the guide. The current in the

walls of the guide is then like current in coils wound on the cores—except that the circuit is completed by displacement current instead of conduction current through the central region of the core.

Consider the current in the side wall at $x = 0$. The magnetic field at the surface of this wall is longitudinal, and therefore the flow of current is parallel to the y axis. The linear current density, denoted by $\mathcal{J}_y$, is, as shown in Art. 2, equal in magnitude to the magnetic-field intensity $\mathcal{H}_z$ at the conductor surface. Hence, by Eq. 45, with $x = 0$,

$$\mathcal{J}_y = -\mathcal{H}_z = -\frac{\mathcal{E}_m \lambda_0}{Z_i \lambda_{co}} \cos(\omega t - \beta z) \tag{46}$$

The negative sign in Eq. 46 arises because positive reference directions for $\mathcal{J}_y$ and $\mathcal{H}_z$ are in the directions of the positive y and z axes, and the right-

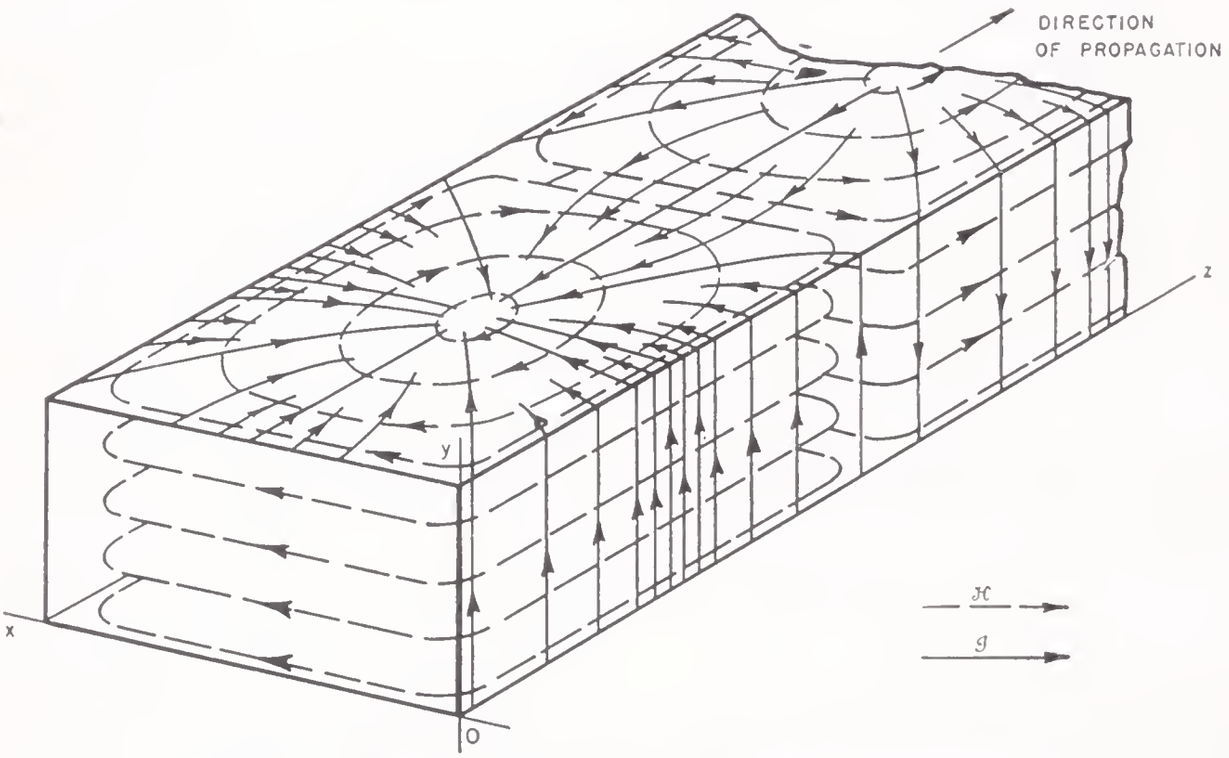


FIG. 15. Current distribution in a rectangular waveguide.

hand rule shows $\mathcal{J}_y$ to be positive (directed upward in Fig. 15) when $\mathcal{H}_z$ is negative (directed to the left along the z axis in Fig. 15). The current density specified by Eq. 46 varies sinusoidally along the z axis, as indicated by the spacing of lines in Fig. 15, and this distribution pattern propagates as a wave in the positive z direction.

The current in the top and bottom walls of the guide may be expressed as the sum of linear current-density components $\mathcal{J}_x$ and $\mathcal{J}_z$ directed parallel to the x and z axes, respectively. The $\mathcal{J}_x$ component in the top wall, for example, is related to the value of $\mathcal{H}_z$ (the longitudinal component of $\mathcal{H}$) at the surface of the top wall, and the $\mathcal{J}_z$ component is related to the component $\mathcal{H}_x$ of magnetic-field intensity. Thus reasoning similar to

that employed in obtaining Eq. 46 yields

$$\mathcal{I}_x = -\mathcal{K}_z = -\left(\frac{\mathcal{E}_m\lambda_0}{Z_i\lambda_{co}}\cos\frac{\pi x}{a}\right)\cos(\omega t - \beta z) \tag{47}$$

$$\mathcal{I}_z = +\mathcal{K}_x = \left(\frac{\mathcal{E}_m\lambda_0}{Z_i\lambda}\sin\frac{\pi x}{a}\right)\cos(\omega t + 90° - \beta z) \tag{48}$$

Equations 47 and 48 together indicate the moving current pattern represented for the top wall in Fig. 15, just as Eqs. 44 and 45 together represent the magnetic-field pattern. The current patterns can be verified by vector addition of the components $\mathcal{I}_x$ and $\mathcal{I}_z$ at various points. Thus, at the edges of the top wall, ($x = 0$ or $x = a$), $\mathcal{I}_z = 0$, and the current is transverse; at $x = a/2$, $\mathcal{I}_x = 0$, and the current is longitudinal. Because of the 90-deg phase relation between $\mathcal{I}_x$ and $\mathcal{I}_z$, the maximum longitudinal current at the center occurs for values of z midway between those for which maximum transverse current occurs at the edges of the top wall.

Many of the properties of waveguides can be understood more readily if the magnetic field and current density are considered in terms of their components (as specified by Eqs. 44, 45, 47, and 48) rather than as total vectors (as indicated in Fig. 15). The electric field, which is directed parallel to the y axis, may be considered in conjunction with the transverse magnetic field and the longitudinal current component as a wave propagating along the length of the guide. This wave is essentially similar to the transverse wave on a parallel-strip line (see Art. 2), but the wavelength is greater, and the electric field, magnetic field, and current density are all greatest at the center of the guide and decrease to zero at the edges. Alternatively, the electric field may be considered in conjunction with the longitudinal magnetic field and the transverse current component as a wave that propagates across the guide and produces a standing-wave pattern between the two side walls. The complete $TE_{1,0}$ wave, therefore, includes both a transverse standing wave and a longitudinal traveling wave. In this sense, a guide carrying the $TE_{1,0}$ wave may be likened to a parallel-strip transmission line with its two conductors joined on each side by a continuous line of quarter-wave stubs.

8. Phase Velocity and Group Velocity. If a rectangular waveguide transmits a continuous, sinusoidal, $TE_{1,0}$ wave, the field pattern in the guide has the configuration represented by Fig. 13. The pattern moves through the guide with a velocity v_p, called the *phase velocity*, which is the product of the frequency and the length of the waves. In terms of the wavelength relation of Eq. 41,

$$v_p = f\lambda = \frac{f\lambda_0}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} = \frac{c}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} \tag{49}$$

For all wavelengths less than λ_{co}, the denominator in Eq. 49 is less than unity; therefore v_p exceeds c, the speed of light. When λ_0 is much less than λ_{co}, v_p is only slightly greater than c, but v_p increases without limit as the source frequency decreases toward the value for cutoff.

A phase velocity greater than the speed of light has no important significance. An analogous phenomenon is the speed at which a wave-shaped water boundary moves along a vertical wall of a tank partly filled with water. When a surface wave strikes the wall at an oblique angle, the speed of the boundary wave along the wall depends upon the direction of approach of the surface wave and increases without limit as normal incidence is approached.

In a radar system, an r-f pulse, or a group of sine-wave cycles, is transmitted by a waveguide. The motion of the field pattern is then somewhat more complicated. Figure 16, which is a plot of electric-field

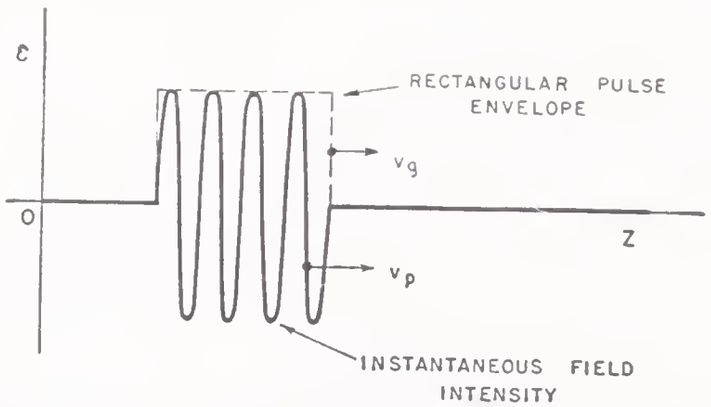


FIG. 16. Rectangular pulse illustrating phase velocity and group velocity.

intensity along the axis of a long section of guide, illustrates the pulse. The individual sinusoidal cycles of the pulse move at the phase velocity v_p, but the rectangular envelope of the pulse moves at a slower speed called the group velocity v_g. Thus the sine-wave cycles appear to move forward through the pulse envelope. As old cycles disappear at the leading edge, new cycles form at the trailing edge.

The different velocities of the pulse-envelope and carrier-frequency waveforms may be explained in terms of a plot of waveguide phase constant β as a function of angular frequency ω. According to Eq. 28, β is equal to $\beta_0 \sin \alpha$, where β_0 (the phase constant of a plane transverse wave) is ω/c, and $\sin \alpha$, from Eq. 41, is $\sqrt{1 - (\lambda_0/\lambda_{co})^2}$. Hence,

$$\beta = \frac{\omega}{c}\sqrt{1 - \left(\frac{\lambda_0}{\lambda_{co}}\right)^2} = \frac{\omega}{c}\sqrt{1 - \left(\frac{\omega_{co}}{\omega}\right)^2} = \frac{\sqrt{\omega^2 - \omega_{co}^2}}{c} \qquad (50)$$

Figure 17 is a graphical representation of this equation. In this figure, ω_c is the angular frequency of the carrier wave, and $\Delta\omega$ represents a frequency band within which lie all the important Fourier components of the pulse. The β, ω curve may be regarded as a curve of phase shift in a

fixed length of guide as a function of frequency, because βd, the total phase shift in a guide section of length d, is proportional to β.

If the β, ω curve were the dashed straight line drawn through the origin in Fig. 17, each Fourier component transmitted by the guide would be delayed by a phase angle proportional to its frequency. All Fourier components would therefore be delayed by the same interval of time, and phase relations between components would remain unchanged. In traveling through the guide, the pulse shape would be perfectly preserved, and the pulse-envelope and carrier-frequency waveforms would travel with the same velocity.

Because the β, ω curve of Fig. 17 differs from a straight line through the origin, the various Fourier components of the pulse are delayed by differ-

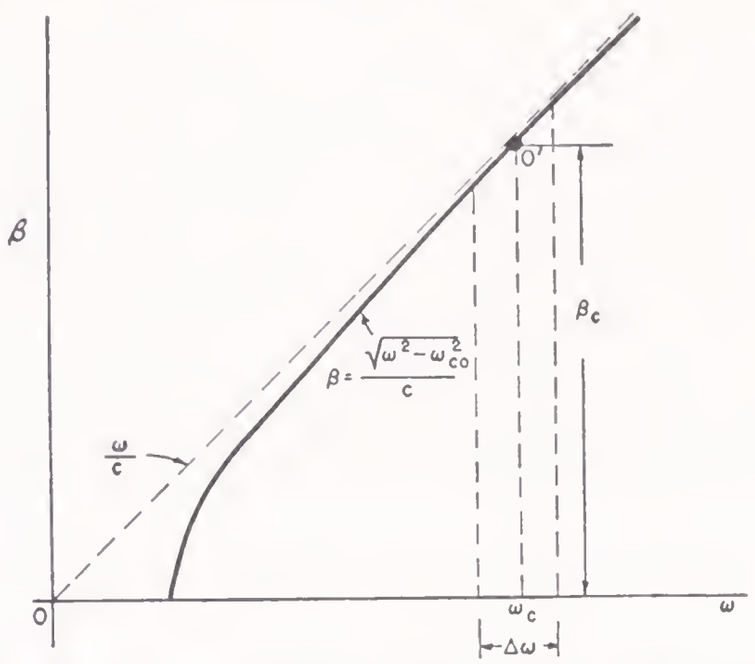


FIG. 17. Waveguide phase constant β as function of angular frequency ω.

ent time intervals. As a result, the shape of the pulse is changed. A special type of pulse distortion occurs, however, because of the fact that within the frequency band of interest $\Delta\omega$, the β, ω curve is essentially a straight line which does *not* pass through the origin when extended. As a result of this linearity, *the shape of the pulse envelope remains unchanged while the carrier-frequency cycles move within the pulse.*

The reason for this type of pulse distortion may be explained in terms of the relation between the responses of carrier-frequency and video circuits discussed in Art. 8 of Chap. VI. The principles developed in that article indicated that the carrier-frequency phase shift is determined by β_c, the value of β at O' in Fig. 17, and the modification of the pulse envelope is the same as that which would occur if the envelope were passed through a circuit having the β, ω curve of Fig. 17 with the origin shifted from O to O'. Since this new β, ω curve is a straight line through the new origin, all the video-frequency Fourier components of the pulse

envelope are delayed by the same time interval. The envelope shape is therefore unaltered.

The phase delay of the carrier frequency in a guide section of length d is $\beta_c d$, and the time delay is $\beta_c d$ divided by ω_c, the angular frequency of the carrier wave, or $\beta_c d/\omega_c$. The phase velocity v_p equals the guide length divided by this time delay and is therefore ω_c/β_c; that is, *phase velocity equals the reciprocal of the slope of a line from O to O′ in Fig.* 17. The phase delay of a video-frequency component of the pulse envelope, evaluated from the β, ω curve with origin at O', is $(\beta - \beta_c)d$; the time delay is $(\beta - \beta_c)d/(\omega - \omega_c)$; and the envelope velocity, or group velocity v_g, is $(\omega - \omega_c)/(\beta - \beta_c)$. Because the β, ω curve is essentially straight within the frequency band $\Delta\omega$, $(\omega - \omega_c)/(\beta - \beta_c)$ has very nearly a constant value equal to the derivative $d\omega/d\beta$ evaluated at O'. Thus *group velocity equals the reciprocal of the slope of the β, ω curve at point O′ in Fig.* 17.

The phase velocity of waves within an air-filled guide always exceeds the speed of light and increases indefinitely as the angular carrier frequency approaches ω_{co}. In contrast, the group velocity is always less than the speed of light and approaches zero as ω tends to ω_{co}. These relations may be inferred from the slope of the line OO' and of the β, ω curve at O' in Fig. 17, for various positions of O'—the reciprocal of the slope of the dashed line representing ω/c corresponds to the speed of light. Alternatively, the value of the phase velocity is given by Eq. 49, and the group velocity may be obtained by differentiating Eq. 50 as follows:

$$\frac{1}{v_g} = \frac{d\beta}{d\omega} = \frac{1}{c\sqrt{1 - (\lambda_0/\lambda_{co})^2}}$$

and

$$v_g = c\sqrt{1 - \left(\frac{\lambda_0}{\lambda_{co}}\right)^2} \tag{51}$$

The factor $\sqrt{1 - (\lambda_0/\lambda_{co})^2}$ is less than unity, and v_g is less than c.

Note, by comparing Eqs. 49 and 51, that v_g is always smaller than c in the same ratio that c is smaller than v_p. The product $v_p v_g$ therefore has the constant value c^2. Curves of v_p/c and v_g/c are plotted in Fig. 18 as a function of λ_0/λ_{co}.

9. Power and Attenuation. The power transmitted by a waveguide may be determined from the Poynting vector $\mathcal{S}$ (see Art. 2) evaluated at points on a cross section of the guide. Unlike the Poynting vector for the parallel-strip line, $\mathcal{S}$ in a waveguide has a component in the transverse plane of the guide as well as a component along the axis of the guide. The electric field in the guide is parallel to the y axis, and therefore the Poynting vector (being perpendicular to both the electric and the magnetic field) has a longitudinal component equal in magnitude to $\mathcal{E}\mathcal{H}_x$ and a

component in the x direction of magnitude $\mathcal{E}\mathcal{H}_z$. The transverse component of $\mathcal{S}$ is associated with the transverse standing-wave aspect of the waveguide fields and specifies the flow of energy required for alternate storage of energy in the electric field at the guide center and in the longitudinal magnetic field near the side walls. The longitudinal component of $\mathcal{S}$ is associated with the traveling-wave aspect of the fields and specifies the power actually transmitted along the guide.

The power transmitted may be expressed in terms of $\mathcal{E}_m$ and $\mathcal{H}_{xm}$, the peak values of the electric-field intensity and transverse magnetic-field intensity, each measured at the center of the guide. By Eqs. 43 and 44,

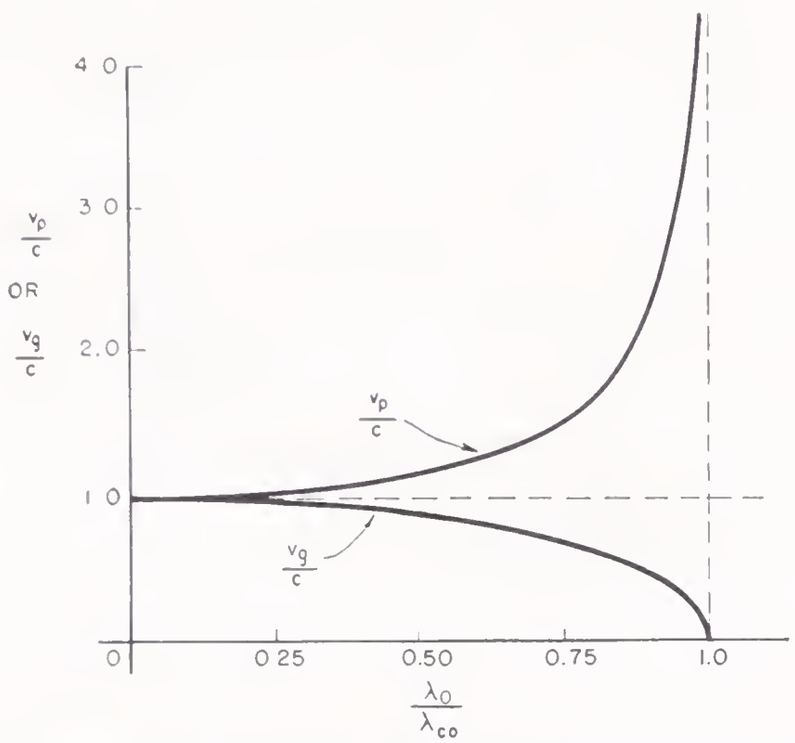


FIG. 18. Variations of phase and group velocity with wavelength λ_0.

the peak values of $\mathcal{E}$ and $\mathcal{H}_x$ at any point of a cross section are $\mathcal{E}_m \sin \pi x/a$ and $\mathcal{H}_{xm} \sin \pi x/a$, and $\mathcal{H}_{xm}$ equals $(\mathcal{E}_m/\mathcal{Z}_i)(\lambda_0/\lambda)$. Because these quantities vary sinusoidally and in phase, the *time average* of the product, or the average power transmitted per unit cross-sectional area, is half the product of peak values. Thus

$$\mathcal{S}_{avg} = \frac{1}{2} \mathcal{E}_m \mathcal{H}_{xm} \sin^2 \frac{\pi x}{a} \tag{52}$$

Equation 52 indicates that the power transmitted per unit area is not the same at all points of the cross section but varies with x from a maximum at the guide center to zero at the side walls.

Consider the element of area shown in Fig. 19 for which the height is b and the width is Δx. If Δx is small, the value of $\mathcal{S}_{avg}$ is essentially constant over the element, and by Eq. 51 the total power transmitted through

the element is

$$\Delta P = \frac{1}{2}\mathcal{E}_m\mathcal{H}_{xm}\sin^2\frac{\pi x}{a}\,b\Delta x$$

The total power transmitted by the guide is the sum of the powers through a large number of such elements that together make up the

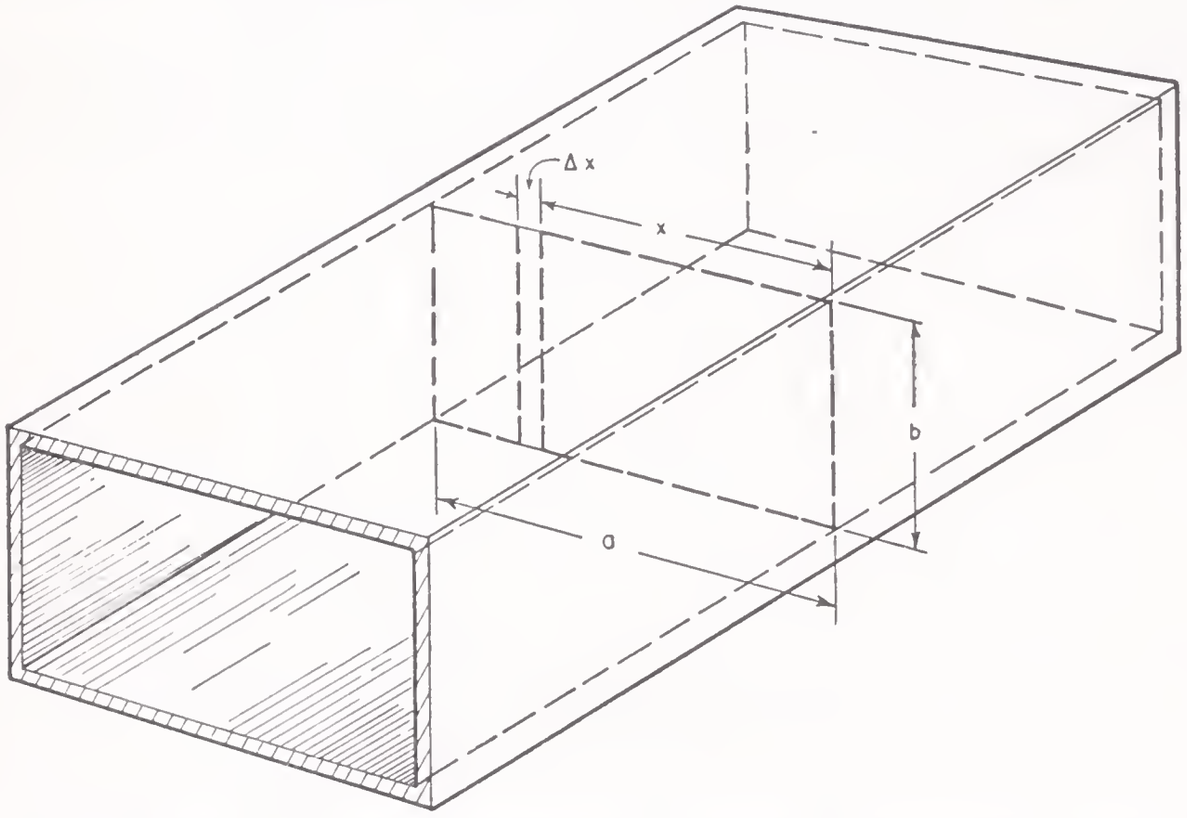


FIG. 19. Element of area of cross section for calculation of transmitted power.

entire cross section. This sum is expressed by the integral

$$P = \int_0^a \frac{1}{2}\mathcal{E}_m\mathcal{H}_{xm}\sin^2\frac{\pi x}{a}\,b\,dx = \frac{b}{2}\mathcal{E}_m\mathcal{H}_{xm}\int_0^a \sin^2\frac{\pi x}{a}\,dx$$

Since $\sin^2 \pi x/a = \frac{1}{2}(1 - \cos 2\pi x/a)$,

$$P = \frac{b}{4}\mathcal{E}_m\mathcal{H}_{xm}\left(\int_0^a dx - \int_0^a \cos\frac{2\pi x}{a}\,dx\right)$$

or

$$P = \tfrac{1}{4}\mathcal{E}_m\mathcal{H}_{xm}ab \tag{53}$$

Observe that Eq. 53 could have been obtained directly by use of the fact that *the average value of the square of a sine wave over a half cycle is one-half.* The peak transmitted power per unit area at the center of the guide is $\mathcal{E}_m\mathcal{H}_{xm}$. To obtain the total average power transmitted, this quantity is multiplied by the cross-sectional area ab, by a factor $\frac{1}{2}$ to account for the sine-squared variation of the power with time, and by another factor $\frac{1}{2}$ to account for the sine-squared variation across the

guide. The result is Eq. 53. Alternate forms of Eq. 53 may be obtained from the relation between $\mathfrak{H}_{xm}$ and $\mathcal{E}_m$,

$$\mathfrak{H}_{xm} = \frac{\mathcal{E}_m}{Z_i}\frac{\lambda_0}{\lambda} = \mathcal{E}_m \frac{\sqrt{1 - (\lambda_0/\lambda_{co})^2}}{Z_i}$$

Because $\mathfrak{H}_{xm}$ is proportional to $\mathcal{E}_m$, P is proportional to $\mathcal{E}_m^2$ or to $\mathfrak{H}_{xm}^2$ if the guide dimensions and operating frequency are fixed.

According to Eq. 53, the transmitted power is the same for every cross section along the guide. This conclusion is the result of the assumption that the guide is free of loss. In actual guides, however, losses do exist, and in consequence the field intensities and the transmitted power attenuate as the wave moves through the guide. The amount of the

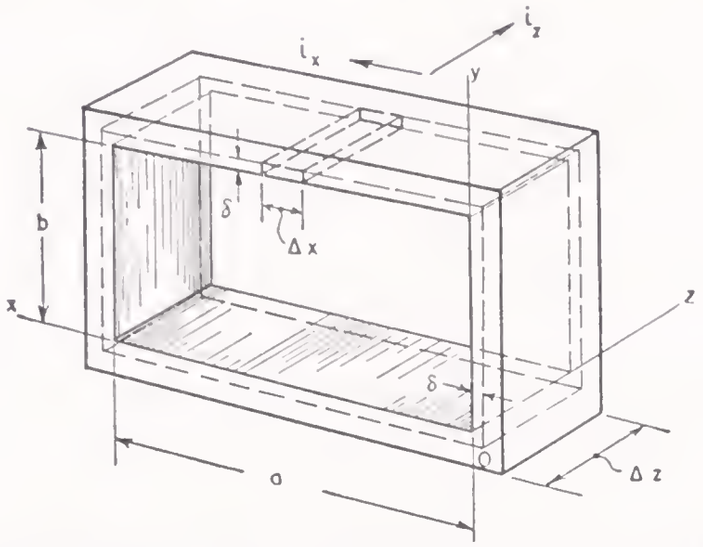


FIG. 20. Short guide section considered in calculation of attenuation.

attenuation is of great importance because it determines the efficiency of operation of the guide.

In a practical, air-filled waveguide the losses in the air dielectric are negligible so that only the loss caused by current flow in the imperfectly conducting guide walls need be considered. Even the guide-wall loss is so small that attenuation is appreciable only over relatively long sections of guide. In a section less than a few wavelengths long, the field patterns are indistinguishable from those of a lossless guide. Because of the smallness of the losses, the resulting attenuation can be calculated by a perturbation method similar to the one employed in calculating transmission-line attenuation (see Art. 13, Chap. VII). The field pattern in a short guide section is assumed to be the same as that in a lossless guide, and from the magnetic field of this pattern the guide-wall currents are determined and the losses calculated.

A short section of guide of length Δz in which the power loss is to be determined is indicated in Fig. 20. Because the resistance of the wall material is not zero, current penetrates somewhat into the walls. As

explained in Art. 15 of Chap. VII, the current density decreases exponentially with distance below the surface, but the actual distribution may be replaced by a current of uniform density in a surface layer of thickness δ. The power loss per unit area of this surface layer may be related to the magnetic field at the conductor surface in the following manner: Consider a small element of the surface layer, for example, the element of thickness δ, length Δz, and width Δx indicated in the figure. As shown in Art. 7, the current in this element comprises (1) a transverse component i_x related to the longitudinal component of magnetic-field intensity by the equation $i_x = \mathcal{H}_z \Delta z$ and (2) a longitudinal component i_z related to the transverse magnetic field by $i_z = \mathcal{H}_x \Delta x$. Because these two current components flow in perpendicular directions, the total power loss in the element at any instant is the sum of the losses occasioned by the two currents acting separately,[1] and is

$$\Delta p = i_x^2 \rho \frac{\Delta x}{\delta \Delta z} + i_z^2 \rho \frac{\Delta z}{\delta \Delta x} = \frac{\rho}{\delta} \mathcal{H}_z^2 \Delta x \Delta z + \frac{\rho}{\delta} \mathcal{H}_x^2 \Delta x \Delta z$$

or

$$\Delta p = \frac{\rho}{\delta} (\mathcal{H}_x^2 + \mathcal{H}_z^2) \Delta x \Delta z \tag{54}$$

where ρ is the resistivity of the wall material and $\rho \Delta x / \delta \Delta z$ and $\rho \Delta z / \delta \Delta x$ are the resistances through which flow the currents i_x and i_z, respectively. Because $\Delta x \Delta z$ is the surface area of the element in which the loss occurs, and because $\mathcal{H}_x^2 + \mathcal{H}_z^2 = \mathcal{H}^2$, the square of the total magnetic-field intensity at the element, Eq. 54 states that $\Delta p / \Delta x \Delta z$ (the instantaneous power loss per unit area of surface) is equal to $(\rho/\delta)\mathcal{H}$. Consequently, the relation of average-power dissipation to the peak magnetic-field intensity $\mathcal{H}_m$ for sinusoidal variation of field intensity and current is

$$\left.\begin{array}{l}\text{Average-power loss per unit} \\ \text{area of conductor surface}\end{array}\right\} = \frac{1}{2} \frac{\rho}{\delta} \mathcal{H}_m^2 \tag{55}$$

The total power loss for the guide section of Fig. 20 may be found from Eq. 55 and from the pattern of the magnetic field for the $TE_{1,0}$ wave described in Art. 7. Let $\mathcal{H}_{xm}$ be the peak transverse magnetic-field intensity at the guide center as in Eq. 53, and let $\mathcal{H}_{zm}$ be the peak longitudinal intensity at the side-wall surface. By Eq. 55, then, the power

[1] This statement is true only because the current directions are perpendicular so that the square of the total current density is the sum of the squares of the component current densities. If, for example, a 2-amp current in a 1-ohm resistor were resolved into two 1-amp currents flowing in the same direction, the total power (4 watts) would not equal the sum (2 watts) of the power dissipations of the component currents acting separately.

loss per unit area is

$$\frac{1}{2}\frac{\rho}{\delta}\,\mathcal{H}_{xm}^2 \qquad \text{for any point on the side walls}$$

and

$$\frac{1}{2}\frac{\rho}{\delta}\left(\mathcal{H}_{xm}^2\sin^2\frac{\pi x}{a}+\mathcal{H}_{zm}^2\cos^2\frac{\pi x}{a}\right) \qquad \text{for any point on the top and bottom walls}$$

The total power loss in the two side walls is found by multiplying the power per unit area by the wall area $2b\Delta z$, and the loss in the top and bottom walls is obtained by first introducing the one-half factor for the average value of the square of a sinusoid and then multiplying the average per-unit-area loss by the wall area $2a\Delta z$. Thus the power loss in the entire guide section is

$$\Delta P = \frac{1}{2}\frac{\rho}{\delta}\mathcal{H}_{zm}^2(2b\Delta z)+\frac{1}{4}\frac{\rho}{\delta}(\mathcal{H}_{xm}^2+\mathcal{H}_{zm}^2)(2a\Delta z)$$

This expression may be simplified by use of the relations

$$\left(\frac{\mathcal{H}_{zm}}{\mathcal{H}_{xm}}\right)^2=\left(\frac{\lambda}{\lambda_{co}}\right)^2 \qquad \text{and} \qquad 1+\left(\frac{\lambda}{\lambda_{co}}\right)^2=\left(\frac{\lambda}{\lambda_0}\right)^2$$

obtained from Eqs. 42, 44, and 45, p. 575, with the result

$$\Delta P = \frac{1}{2}\frac{\rho b}{\delta}\mathcal{H}_{xm}^2\left[2\left(\frac{\lambda}{\lambda_{co}}\right)^2+\frac{a}{b}\left(\frac{\lambda}{\lambda_0}\right)^2\right]\Delta z$$

or

$$\Delta P = \frac{1}{2}\frac{\rho b}{\delta}\left(\frac{\lambda}{\lambda_0}\right)^2\mathcal{H}_{xm}^2\left[2\left(\frac{\lambda_0}{\lambda_{co}}\right)^2+\frac{a}{b}\right]\Delta z \tag{56}$$

Equation 56 indicates that the power loss in a section of guide is proportional to $\mathcal{H}_{xm}^2$, and as P, the power transmitted, is also proportional to $\mathcal{H}_{xm}^2$, the power loss is proportional to the power transmitted. This same relation of power lost to power transmitted occurs in transmission lines (see Art. 13, of Chap. VII), and as a result the field intensities and the power transmitted decrease exponentially with distance along the line or guide. The rate of decrease is measured by the *attenuation constant* α, and the relation

$$P = P_0\epsilon^{-2\alpha z} \tag{57}$$

applies, where P_0 is the value of P at a reference cross section for which $z = 0$. From Eq. 59, p. 517, the attenuation constant is

$$\alpha = \frac{1}{2}\frac{-dP/dz}{P} = \frac{1}{2}\frac{\text{power loss per unit length}}{\text{power transmitted}} \tag{58}$$

The value of α can be determined from Eq. 58 if for power loss per unit length the value of $\Delta P/\Delta z$ from Eq. 56 is employed and if for power transmitted the value of P from Eq. 53 (with $\mathcal{E}_m$ replaced in terms of $\mathcal{H}_{xm}$) is used. The result is

$$\alpha = \frac{1}{2} \frac{\frac{1}{2} \frac{\rho b}{\delta} \left(\frac{\lambda}{\lambda_0}\right)^2 \mathcal{H}_{xm}^2 \left[2\left(\frac{\lambda_0}{\lambda_{co}}\right)^2 + \frac{a}{b}\right]}{\frac{1}{4} \mathcal{H}_{xm}^2 Z_i \frac{\lambda}{\lambda_0} ab}$$

or

$$\alpha = \frac{\rho}{Z_i \delta a} \frac{a/b + 2(\lambda_0/\lambda_{co})^2}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} \tag{59}$$

A more directly useful form for Eq. 59 is obtained as follows: The intrinsic impedance Z_i of air or space is $\sqrt{\mu_0/\varepsilon_0}$, and the skin depth δ for nonferromagnetic conductors is $\sqrt{\rho/\pi\mu_0 f}$ or $\sqrt{\rho\lambda_0/\pi\mu_0 c}$. Thus the attenuation relation becomes

$$\alpha = \sqrt{\pi\varepsilon_0 c} \frac{\sqrt{\rho}}{a} \frac{1}{\sqrt{\lambda_0}} \frac{a/b + 2(\lambda_0/\lambda_{co})^2}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}}$$

Multiplication by $\sqrt{\lambda_{co}/\lambda_{co}} = \sqrt{\lambda_{co}/2a}$ yields

$$\alpha = \sqrt{\frac{\pi\varepsilon_0 c}{2}} \frac{\sqrt{\rho}}{a^{3/2}} \sqrt{\frac{\lambda_{co}}{\lambda_0}} \frac{a/b + 2(\lambda_0/\lambda_{co})^2}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} \tag{60}$$

In rationalized mks units a and b in Eq. 60 are in meters, ρ is in ohm-meters, c in meters per second, ε_0 in farads per meter, and α in nepers per meter of guide length.

Evaluation of $\sqrt{\pi\varepsilon_0 c/2}$ yields 0.0647 ohm$^{-1/2}$. Furthermore, 1 neper is 8.686 db (see Art. 13, Chap. VII). Hence, for an air-filled guide the attenuation is

$$\alpha = 0.561 \frac{\sqrt{\rho}}{a^{3/2}} \sqrt{\frac{\lambda_{co}}{\lambda_0}} \frac{a/b + 2(\lambda_0/\lambda_{co})^2}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} \quad \text{db per m} \tag{61}$$

In Fig. 21 the quantity $\alpha a^{3/2}/\sqrt{\rho}$ is plotted as a function of $\lambda_{co}/\lambda_0 = f/f_{co}$. The shapes of the curves result from the combined effect of the three factors involving wavelength in Eq. 61. The curves indicate that the attenuation is high near cutoff and again at frequencies far above cutoff. In the derivation of Eq. 61, the assumption was made that losses were small. This assumption is not justified at frequencies very close to cutoff where the attenuation given by Eq. 61 becomes exceedingly large.

This equation should be employed only at frequencies for which it specifies relatively small values of attenuation, and the curves of Fig. 21 are plotted in such a range.

Figure 21 indicates that operation at a frequency well above cutoff is desirable to obtain low-loss transmission. Excessively high frequencies relative to the cutoff frequency should be avoided, however, not only

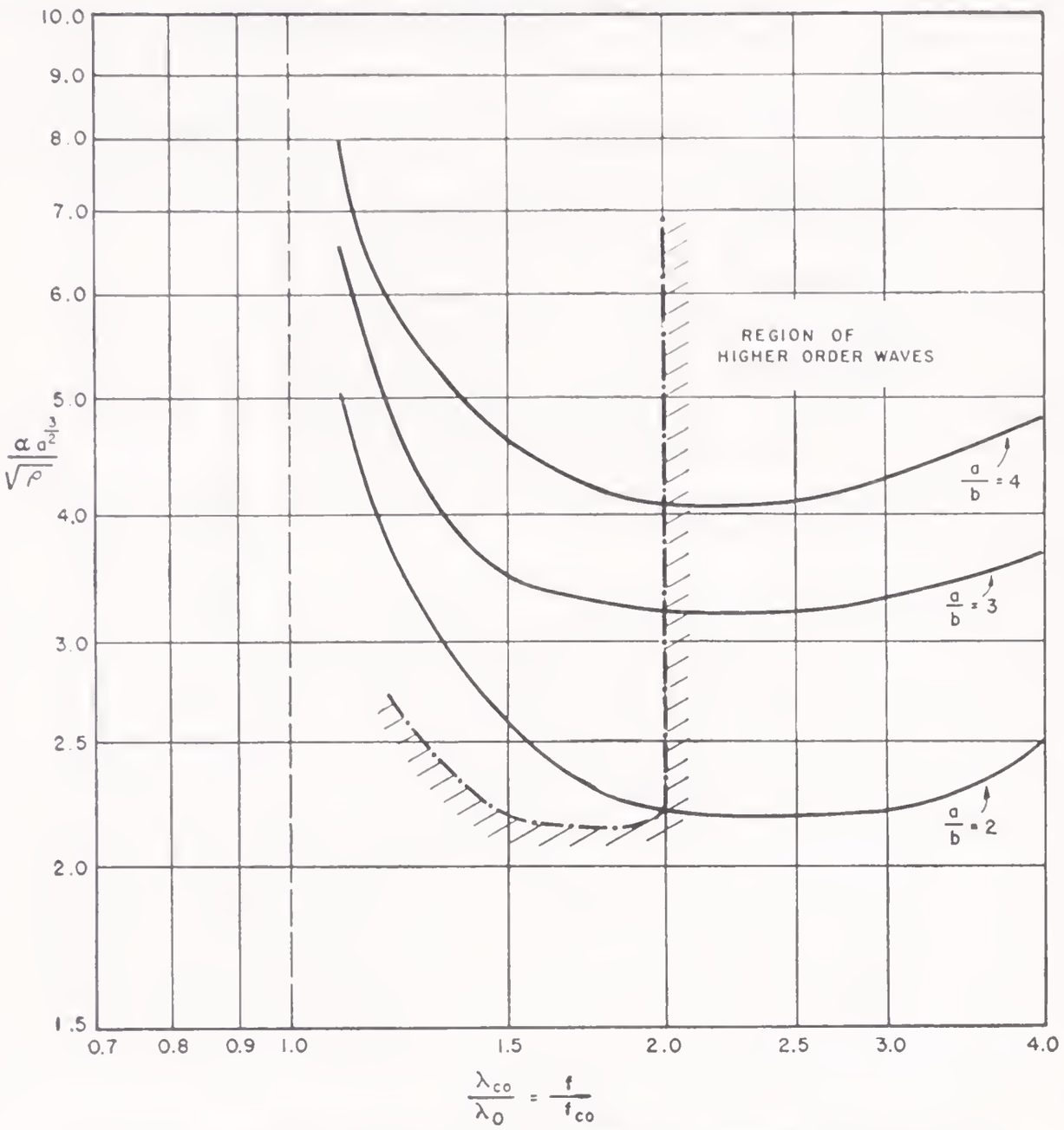


FIG. 21. Attenuation curves for $TE_{1,0}$ wave in a rectangular waveguide.

because of the increased attenuation but also because higher order waves can be transmitted at these frequencies. As explained in Arts. 10 and 11, higher order waves are not transmitted if the transverse-wave wavelength λ_0 is greater than a and greater than $2b$. Thus the operating frequency should lie to the left of the dash-dot line in Fig. 21.

10. Higher Order Wave Types in Rectangular Waveguides. Under certain conditions, types of transverse-electric waves other than the $TE_{1,0}$ wave described in the previous articles and also transverse-mag-

netic waves may be propagated within a rectangular waveguide. If, for example, the spacing of the reflecting plates in Fig. 11 is made equal to $\lambda_0/\cos\alpha$ instead of $\lambda_0/(2\cos\alpha)$, two half-cycle variations of electric-field intensity occur along the larger transverse dimension of the guide, and the wave propagated, called a $TE_{2,0}$ wave, has the pattern of Fig. 22*A*. In more general terms, a type $TE_{m,0}$ wave may be set up within the guide when the side-wall spacing is $m\lambda_0/(2\cos\alpha)$, where m signifies the number of half-period variations in the electric-field intensity along the a dimension of the guide.

Another group of transverse-electric wave types results when rays of the source waves are directed obliquely toward the top or bottom wall of the guide and polarization is along the larger transverse dimension. The component waves within the guide, instead of being reflected from the side walls, are reflected from the top and bottom walls, and the resultant $TE_{0,n}$ wave within the guide has a field pattern rotated 90 deg from that of a $TE_{m,0}$ wave, as indicated in Fig. 22*B*. The subscript n in the type designation denotes the number of half-period variations of the electric-field intensity along the smaller transverse dimension.

If a ray of a source wave is directed obliquely into a corner of a rectangular guide, reflection occurs from all four walls. If the electric-field lines of the source wave lie along intersections of the wavefront with transverse planes of the guide, the wave transmitted in the guide is *transverse electric*. The electric-field intensity of the guided wave varies along both transverse dimensions of the guide, and the wave is designated as a $TE_{m,n}$ wave, where m and n represent the number of half-period variations of electric-field intensity along the a and b dimensions, respectively. The field pattern of a $TE_{1,1}$ wave is illustrated in Fig. 22*C*.

If the source wave directed obliquely into a corner of the guide is polarized so that the magnetic-field lines of the source lie in a transverse plane of the guide, the resultant wave within the guide is *transverse magnetic* and is designated as a $TM_{m,n}$ wave. The electric-field intensity of $TM_{m,n}$ waves varies along both transverse dimensions of the guide, and m and n have the same significance as for transverse-electric waves. The field pattern of a $TM_{1,1}$ wave is illustrated in Fig. 22*D*.

The field configuration of $TE_{1,0}$, $TE_{1,1}$, and $TM_{1,1}$ waves may be regarded as basic configurations because the patterns of all the higher order waves may be formed from these three. For example, a $TE_{2,0}$ pattern may be obtained by placing two small guides, each carrying a $TE_{1,0}$ wave, adjacent to each other. Removal of the common walls yields the $TE_{2,0}$ pattern. If neither m nor n is zero, the pattern of a $TE_{m,n}$ wave may be formed by stacking a number of identical small guides in a close rectangular array m guides wide and n guides high. Each of the small guides carries a $TE_{1,1}$ wave, and waves in adjacent

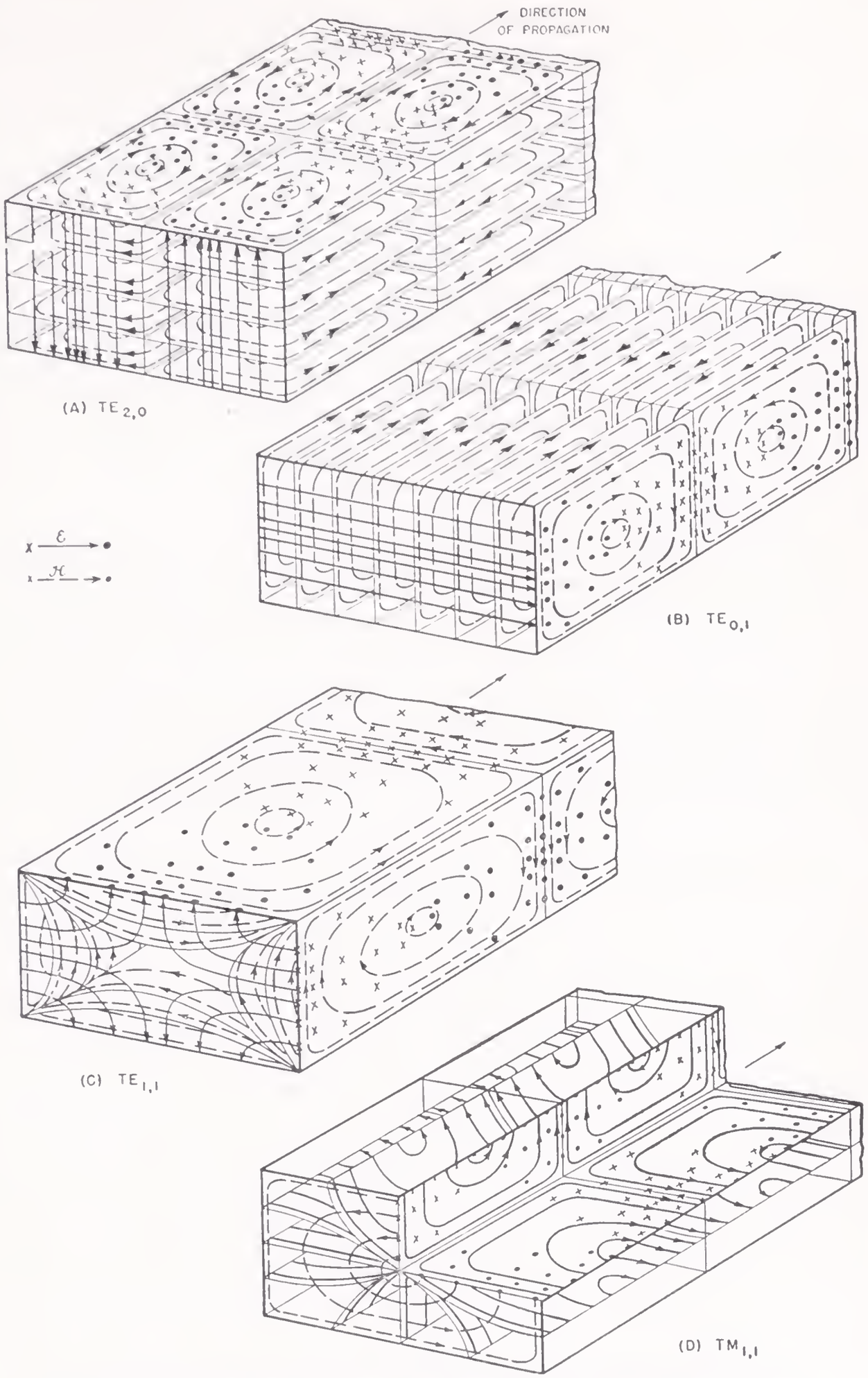


FIG. 22. Examples of higher order wave types for rectangular guides.

guides are in phase opposition. Removal of the inside walls of the array yields the pattern of the $TE_{m,n}$ wave. Patterns of $TM_{m,n}$ waves may be built up in a similar way if each small guide carries a $TM_{1,1}$ wave.

It may be shown that associated with each possible wave type is a cutoff wavelength related to the cross-sectional dimensions of the guide in the following manner:

$$\lambda_{co} = \frac{1}{\sqrt{(m/2a)^2 + (n/2b)^2}} \tag{62}$$

For either transverse-electric or transverse-magnetic waves, m and n may have any positive integral values, 1, 2, 3, ... For transverse-electric waves only, either m or n may also be zero. Thus, in a guide in which a is greater than b, the $TE_{1,0}$ wave has the longest cutoff wavelength and for this reason is called the *dominant wave*. Setting $m = 1$ and $n = 0$ in Eq. 62 yields the cutoff wavelength of the dominant wave $\lambda_{co} = 2a$ in agreement with the value given in Eq. 40, p. 574.

The length of higher order waves along the axis of a guide may be obtained from the guide-wavelength relation for a $TE_{1,0}$ wave (Eq. 41, p. 574), provided in the relation λ_{co} is the cutoff wavelength of the higher order wave. Since λ_{co} is shorter for the higher order waves than for the dominant wave, the higher order waves have the longer guide wavelengths. Similarly, each higher order wave has its own phase and group velocity which may be obtained from the relationship for the $TE_{1,0}$-wave phase and group velocities (Eq. 49 and 51) through use of the appropriate value of λ_{co}.

11. Rectangular-waveguide Dimensions. The dimensions of rectangular waveguides for radar use are usually chosen so that only the dominant wave can be transmitted. Any higher order waves set up within the guide are then attenuated rapidly in the vicinity of their source. Higher order waves may originate at discontinuities (such as bends, twists, matching devices, and transmit-receive devices) which are always present in any practical installation. Elimination of higher order waves is desirable because of the extreme difficulty of designing guide terminations and matching devices to operate satisfactorily with respect to more than a single wave type. To permit propagation of a $TE_{1,0}$ wave, the guide dimension a must exceed $\lambda_0/2$. To prevent propagation of the $TE_{2,0}$ and the higher order $TE_{m,0}$ waves, a must be less than λ_0. The b dimension of the guide must be less than $\lambda_0/2$ in order to prevent propagation of $TE_{0,1}$ and higher order $TE_{0,n}$ waves. Choosing a and b so as to prevent $TE_{2,0}$ and $TE_{0,1}$ waves also ensures that other higher order transverse-electric and transverse-magnetic waves will not be propagated. Thus the cross-sectional dimensions should satisfy the following inequalities:

$$\frac{\lambda_0}{2} < a < \lambda_0$$

$$0 < b < \frac{\lambda_0}{2}$$

Factors that affect the choice of guide dimensions within the limitations of the above inequalities are the attenuation and the amount of power that the guide must transmit. From the attenuation curves of Fig. 21, it may be seen that, in order to keep the attenuation low, a should be chosen so that operation is well above the cutoff frequency and b should be made as large as possible. The maximum power capacity may be determined from Eq. 53, p. 582, expressed in the following form:

$$P = \frac{1}{4}\frac{\mathcal{E}_m^2}{Z_i}\sqrt{1 - \left(\frac{\lambda_0}{\lambda_{co}}\right)^2}\, ab = \frac{1}{4}\frac{\mathcal{E}_m^2}{Z_i}\, b\sqrt{a^2 - (\lambda_0/2)^2} \qquad (63)$$

Thus the power capacity for any operating wavelength λ_0 depends upon $\mathcal{E}_m$, a, and b. The maximum allowable value of $\mathcal{E}_m$ is fixed by the nature of the dielectric, and thus both a and b should each be made as large as possible if higher power capacity is needed. In practice, b is usually made about half a.

A list of Army-Navy standard-size rectangular waveguides, together with some of their properties, is given in Table 1. The attenuations quoted are for brass-walled guides, and the power capacities are based upon a peak field strength of 3 Mv per m and apply only if no standing waves (see Art. 12, Chap. VII) are present in the guide.

TABLE 1. ARMY-NAVY STANDARD-SIZE BRASS RECTANGULAR WAVEGUIDES

Type No.	*Outside dimensions, in.*	*Inside dimensions*				*Cutoff wavelength for $TE_{1,0}$ wave, cm*	*Range of operating wavelengths, cm*	*Attenuation in operating range, db/m*	*Maximum power capacity at $0.7\lambda_{co}$, Mw*
		a		*b*					
		In.	*Cm*	*In.*	*Cm*				
RG-48/U	3 × 1½	2.84	7.22	1.34	3.40	14.44	7.6–11.8	0.034–0.048	10.9
RG-49/U	2 × 1	1.872	4.76	0.872	2.22	9.52	5.0–7.6	0.064–0.086	4.5
RG-50/U	1½ × ¾	1.372	3.48	0.622	1.58	6.96	3.8–5.7	0.104–0.146	2.35
RG-51/U	1¼ × ⅝	1.122	2.86	0.497	1.26	5.72	3.0–4.7	0.152–0.225	1.54
RG-52/U	1 × ½	0.900	2.29	0.400	1.02	4.58	2.4–3.7	0.199–0.278	1.00
RG-53/U	½ × ¼	0.420	1.07	0.170	0.43	2.14	1.1–1.7	0.66 –0.93	0.20
RG-69/U	6.66 × 3.41	6.50	16.52	3.25	8.26	33.0	17.5–26.4	0.092×10^{-4}–0.135×10^{-4}	58.4

12. Waveguide Fields for Frequencies below Cutoff. As explained in Art. 7, $TE_{1,0}$ waves can travel through an air-filled rectangular wave-

guide of width a, provided the frequency of the waves exceeds the cutoff frequency $c/2a$. Waves of lower frequency cannot be propagated in the guide. This statement does not mean that low-frequency fields can never exist inside a guide; rather, low-frequency fields, like static fields, can enter the open end of a guide, but only to the extent of a fringing effect. The fringing fields become negligible at relatively short distances from their source. In contrast, high-frequency fields may be *propagated* long distances through the guide with definite phase velocity and little attenuation.

An example related to the $TE_{1,0}$ wave of a rectangular guide may help to explain the fringing fields and waveguide operation below the cutoff frequency. In Fig. 23*A*, a source of low-frequency alternating current is connected to the open end of a rectangular waveguide at the mid-points of

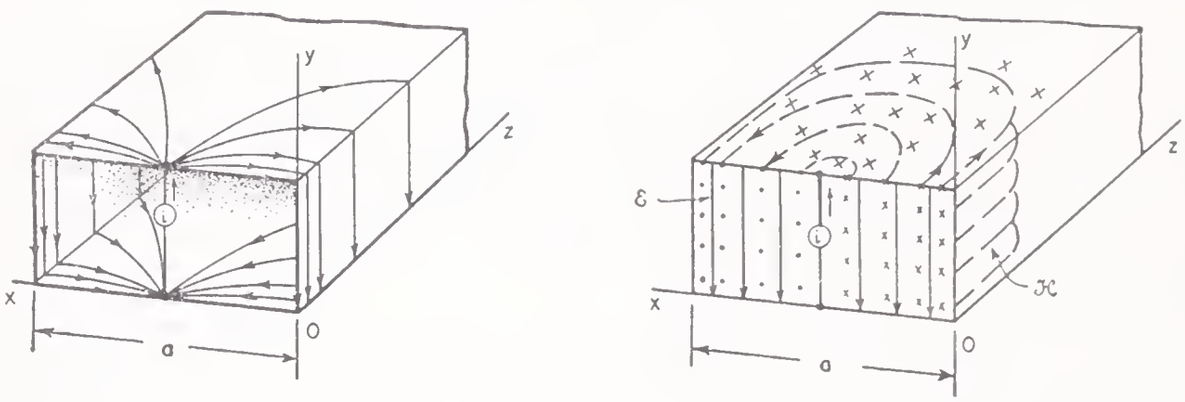


FIG. 23. Fields in a waveguide connected to a low-frequency current source.

the top and bottom walls. The paths of current through the waveguide conductors are indicated roughly by the curved arrows. Lines of magnetic-field intensity encircle the current source, and these are represented by the dashed-line arrows marked $\mathcal{H}$ in Fig. 23*B*. If the resistance of the guide walls is ignored, the voltage at the source terminals will be only the voltage induced by the changing magnetic flux, and the load impedance for the source will be a small inductance. For very low frequencies, the small reactance of this inductance produces a negligible voltage drop, and the corresponding electric field can be ignored.

Increasing the frequency of the source current increases the reactance of the inductance, increases the voltage at the source terminals, and therefore produces an appreciable electric field between top and bottom walls of the guide, as indicated by the solid-line arrows in Fig. 23*B*. The time variation of the electric-field intensity is in phase with the voltage variations and therefore 90 deg out of phase with the magnetic-field intensity variations. The combination of electric and magnetic fields represented in this figure is one of the fringing-field patterns that may occur in a rectangular waveguide at frequencies below cutoff. It is very nearly

the pattern of an *attenuated* $TE_{1,0}$ *wave* (that is, a $TE_{1,0}$ wave at a frequency below cutoff) and would be exactly the field pattern of this wave if the current source were distributed across the end of the guide in such a way as to produce a sinusoidal distribution of the electric-field intensity.

The penetration of the fields of Fig. 23*B* into the guide increases with increase of the source frequency. The displacement current, or rate of change of electric flux, associated with the changing electric field leads the voltage at the source terminals by 90 deg. The conduction currents, in phase with the magnetic-field intensity, lag the voltage by 90 deg. Thus the transverse currents indicated in Fig. 23*A* may be considered partly offset by the displacement current, leaving a greater portion of the source current to flow longitudinally and set up magnetic fields at positions farther into the guide.

If the frequency is sufficiently increased, a resonant condition is reached in which the transverse conduction currents are completely offset by displacement current, and the fields extend at constant amplitude indefinitely far into the guide. In this resonant condition, the guide may be thought of as two very broad parallel-strip lines, each a quarter-wave long (corresponding to $a/2$) and terminated in a short circuit, connected to the current source in parallel. Thus resonance corresponds to the cutoff frequency, and the fields correspond to two transverse waves traveling back and forth between the side walls.

At frequencies above cutoff, a marked change from the field patterns of Fig. 23*B* to those of Fig. 12, p. 572, occurs, and this change can be explained in terms of transverse waves moving diagonally through the guide (see Arts. 6 and 7). Note that the electric-field-intensity pattern in any transverse plane is the same above and below cutoff, but the single set of loops of the magnetic lines obtained at low frequencies breaks into a series of such loops at half-wave intervals, and the position of the loops shifts so that the maximum transverse magnetic-field intensity occurs in the same place as the maximum electric-field intensity.

Associated with the change of field patterns at cutoff is a change in energy relationships. Above cutoff, the electric-field intensity $\mathcal{E}$ and the transverse magnetic-field intensity $\mathcal{H}_x$ are in phase, and the Poynting vector indicates the flow of energy through the guide (see Art. 9). Below cutoff, both $\mathcal{H}_x$ and the longitudinal component $\mathcal{H}_z$ are 90 deg out of phase with $\mathcal{E}$, and the Poynting vector indicates no net transfer of energy through any area inside the guide over an integral number of cycles. Thus propagated waves provide a transfer of energy or power transmission. Attenuated waves represent only stored energy. In general, the appearance of attenuated waves in a guide is associated with the effect of a reactive element because of the stored energy. For the $TE_{1,0}$ waves, the reactive element is always inductive because the net stored energy is

magnetic. (At very low frequencies, only the magnetic field need be considered. At higher frequencies, the electric field is important but stores less energy than the magnetic field for all frequencies below the resonance corresponding to cutoff.)

Attenuated $TE_{1,0}$ waves in a rectangular guide are of interest in a few special circumstances. For example, the fields of these waves are possible sources of feedback in i-f amplifiers (see Art. 21, Chap. VI), and waveguides operated below the cutoff frequency are sometimes used as attenuators in r-f systems (see Art. 8, Chap. XII). Attenuated waves of higher order are of much greater importance, however, because they appear in normal waveguide operation. Many devices in waveguides generate the higher order waves, but because of the choice of waveguide dimensions (Art. 11), the fields of these waves are confined to the vicinity of the sources.

The general appearance of the fields of any attenuated transverse-electric wave is suggested by the relation of the fields of Fig. 23*B* to the fields of a propagated $TE_{1,0}$ wave. The electric field is transverse and has the same pattern as in a propagated wave. The magnetic field is obtained from a single half-wave section of the pattern in a propagated wave, the phase of the transverse-magnetic-field intensity being shifted 90 deg. All transverse-electric waves have the same general properties described for the $TE_{1,0}$ waves; the net stored energy is magnetic at all frequencies below cutoff, and appearance of an attenuated transverse-electric wave in a guide is associated with the effect of an inductance.

The same statements apply to attenuated transverse-magnetic waves, except that the roles of electric and magnetic fields are reversed. The magnetic field, in the transverse plane, has the same pattern as it does in a propagated wave. The stored energy is almost entirely in the electric field at very low frequency, and the net stored energy is electric at any frequency below cutoff. Thus an attenuated transverse-magnetic field is associated with a capacitance. This effect is illustrated for the $TM_{1,1}$ wave by the diagram of Fig. 24, in which a voltage e is applied between the guide wall and a point at the center of the input cross section. For very low frequencies of the applied voltage, the only appreciable field is the electric field represented by the solid-line arrows, and the load for the voltage source is a capacitance. At higher frequencies, magnetic-field lines (indicated by dashed-line arrows) encircle the displacement current associated with the changing electric field.

The changes that occur from one cross section to another as a wave travels through a guide may be expressed in terms of a complex-number propagation constant (see Art. 13, Chap. VII). For a propagated wave, the real part of the propagation constant, the attenuation constant, is zero (except for the effect of power loss) and the imaginary part, the

phase constant, has the value β. Thus the wave travels without loss of amplitude but with progressive increase in phase delay. For an attenuated wave, the reverse occurs. The intensity of the fields decreases exponentially with distance, and the phase does not change. Thus the phase constant is zero, and the attenuation constant is greater than zero.

The value of the attenuation constant for any attenuated wave is related very simply to the phase constant of the corresponding propagated wave. The propagation constant for any wave is $0 + j(\omega/c)\sqrt{1 - (\lambda_0/\lambda_{co})^2}$.

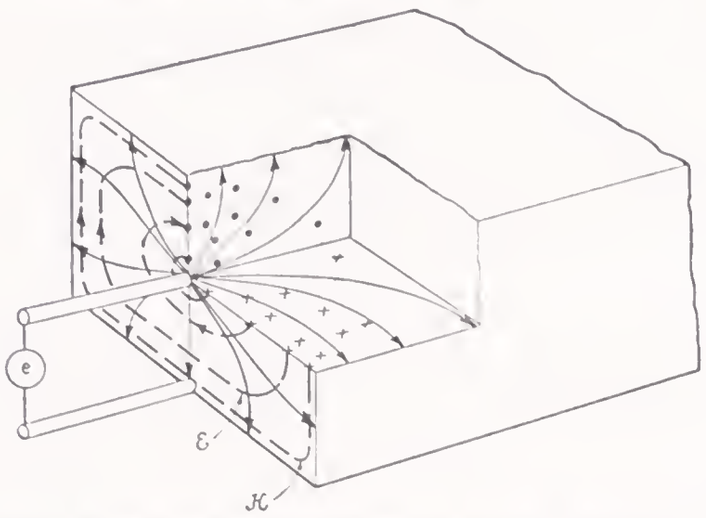


FIG. 24. Fields in a waveguide connected to a low-frequency voltage source.

(This value is obtained from the phase constant β of a $TE_{1,0}$ wave given by Eq. 50, p. 578. It applies also to higher order waves, as explained in Art. 10.) In slightly different form, the propagation constant is

$$j\frac{\omega}{c}\sqrt{1 - \left(\frac{\lambda_0}{\lambda_{co}}\right)^2} = \frac{\omega}{c}\sqrt{(-1)\left[1 - \left(\frac{\lambda_0}{\lambda_{co}}\right)^2\right]} = \frac{\omega}{c}\sqrt{\left(\frac{\lambda_0}{\lambda_{co}}\right)^2 - 1} \quad (64)$$

and thus is a real number for frequencies below cutoff ($\lambda_0 > \lambda_{co}$). This real number is the attenuation constant[1] α of the low-frequency fields. Thus, for attenuated waves,

$$\alpha = \frac{\omega}{c}\sqrt{\left(\frac{\lambda_0}{\lambda_{co}}\right)^2 - 1} = \frac{2\pi}{\lambda_0}\sqrt{\left(\frac{\lambda_0}{\lambda_{co}}\right)^2 - 1} = \pi\sqrt{\frac{1}{a^2} - \left(\frac{2}{\lambda_0}\right)^2} \quad (65)$$

At very low frequencies, $(2/\lambda_0)^2$ becomes negligible relative to $1/a^2$ and

$$\alpha = \frac{\pi}{a} \qquad \text{for } \lambda_0 \gg 2a \quad (66)$$

[1] That the propagation constant derived in terms of above-cutoff fields (see Arts. 5, 6, and 7) should apply to below-cutoff fields is not at all obvious. The propagation constant can, however, be derived in a different way that is valid for all frequencies. The alternate derivation requires the solution of the electromagnetic-wave equation for boundary conditions appropriate to waveguides and may be found in many texts. See, for example, J. C. SLATER, *Microwave Transmission*, 130–138 (McGraw-Hill Book Company, Inc., New York, 1942).

At frequencies far below cutoff, therefore, the intensities of the fringing fields of attenuated waves are reduced by π nepers (that is, multiplied by the factor $\epsilon^{-\pi} = 0.0476$) for each distance a that the point of observation is moved away from the source of the waves. This attenuation may be expressed also as 26.5 db per a dimension.

Observe carefully the differences between the attenuation specified by Eqs. 65 and 66 and the attenuation that occurs above the cutoff frequency because of the resistance of the guide walls. First, the attenuation below cutoff is much greater in magnitude than that due to losses in useful waveguides. Second, the below-cutoff attenuation is *not* the result of a change of electromagnetic energy into heat. At low frequencies, the electric and magnetic fields are 90 deg out of phase, so that no power is transmitted. The guide fails to deliver power to a load, not because the power is lost in transit, but because the guide does not absorb power from the source.

13. Waveguide Impedances. In the preceding articles, the propagation of waves through guides has been studied with little regard to the

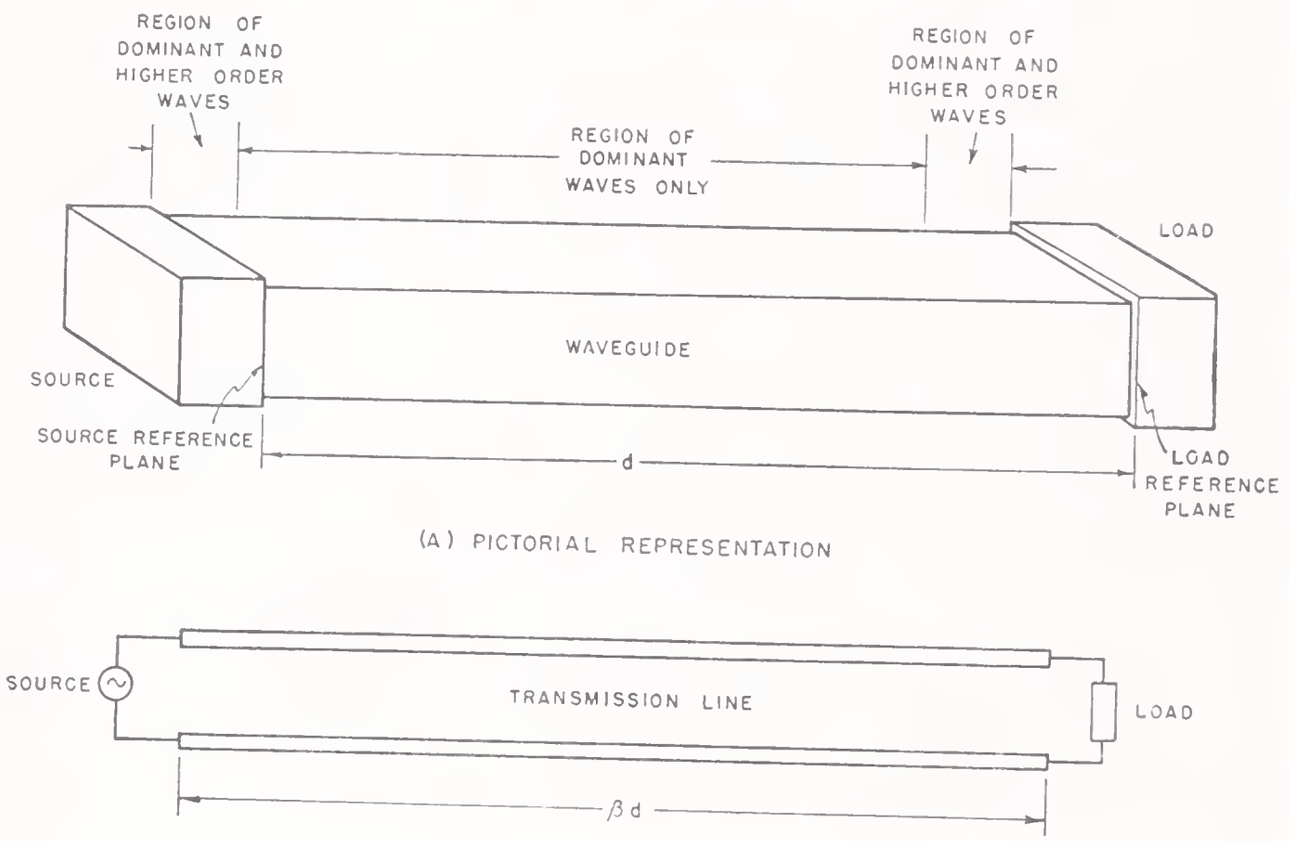


FIG. 25. Operation of waveguide with source and load.

devices used to launch the waves at one end of the guide and receive them at the other. The relation of terminating devices and auxiliary equipment to waveguides is very important. The present article begins the study of such relations by a consideration of the source-guide-load system in Fig. 25*A*. The frequency of the source in this figure is assumed to be

chosen so that dominant waves can be transmitted through the guide, whereas higher order waves are attenuated to negligible amplitudes within a short distance from their points of origin.

Near the input of a waveguide, the configuration of electric and magnetic fields depends largely upon the geometry of the source and ordinarily is very different from the field pattern of a $TE_{1,0}$ wave. The field configuration may, however, be thought of as the sum of $TE_{1,0}$ and higher order waves—much as an arbitrary function of time is considered the sum of its Fourier components. Because of the conditions specified for the source frequency, the higher order waves are restricted to the input region of the guide, and the dominant wave travels through the guide.

The dominant wave travels to the load, and reflections of portions of the wavefront from the various boundary surfaces of the load device cause a resultant field pattern near the load that may be very different from that of the incident $TE_{1,0}$ wave. Ordinarily, two $TE_{1,0}$ waves—an incident wave and a reflected wave—and a variety of higher order waves must be summed to obtain the pattern actually present. The load is thus equivalent to a generator of higher order waves that become negligible at a short distance from the load and of a $TE_{1,0}$ wave that travels from load to source. As suggested by Fig. 25*A*, therefore, the fields in the guide are the sum of two $TE_{1,0}$ waves traveling in opposite directions, plus higher order waves in the terminating regions.

The dominant waves are responsible for the transfer of energy from source to load. Therefore, for most purposes, the higher order waves may be ignored and the waveguide system studied in terms of the relation of source and load to the dominant waves. With respect to dominant waves, the circuit of Fig. 25*B* may be considered an equivalent circuit for the waveguide system. The forward and return waves on the transmission line are analogous to the two dominant waves in the guide.

In Chap. VII, oppositely directed transmission-line waves were found to combine to yield a standing-wave pattern. In exactly the same way, the dominant waves in the guide combine to yield standing waves of electric- and magnetic-field intensities. When there is no reflected wave, the intensities of the electric and magnetic fields have the same rms value at any cross section along the guide (except, of course, for the attenuation caused by losses). With a reflected wave present, the maxima of electric-field intensity occur at intervals of a half wavelength, where the electric fields of the two component waves are in phase. Midway between the electric-field maxima are the minima of the electric field and the maxima of the transverse magnetic field. The physical distance between successive maxima or minima is half the guide wavelength λ (which exceeds λ_0, the length of transverse waves—see Eq. 41, p. 574) and is therefore greater than the corresponding distance for transmission-line waves of

the same frequency. Correspondingly, the length of the transmission line in Fig. 25*B* must be less than the physical length of the guide it represents in order that the lengths in electrical degrees may be equal. Standing-wave patterns in guides may be measured in sections of guide free of higher order waves by techniques very similar to those employed with coaxial lines. Devices for these measurements are described in Art. 18, Chap. XII.

If the standing-wave ratio is unity (that is, if the load is adjusted so that only the dominant wave traveling from source to load is present), then the electric-field intensity $\mathcal{E}$ and transverse magnetic-field intensity $\mathcal{H}_x$ in regions free of higher order waves have the values given by Eqs. 43 and 44, p. 575:

$$\mathcal{E} = \left(\mathcal{E}_m \sin \frac{\pi x}{a}\right) \cos (\omega t + 90^\circ - \beta z)$$

$$\mathcal{H}_x = \left(\frac{\mathcal{E}_m \lambda_0}{\mathcal{Z}_i \lambda} \sin \frac{\pi x}{a}\right) \cos (\omega t + 90^\circ - \beta z)$$

Throughout any cross section, therefore, the variations of $\mathcal{E}$ and $\mathcal{H}_x$ occur in the same phase, and the ratio of $\mathcal{E}$ to $\mathcal{H}_x$ has the constant value

$$\mathcal{Z}_c = \frac{\mathcal{E}}{\mathcal{H}_x} = \mathcal{Z}_i \frac{\lambda}{\lambda_0} = \mathcal{Z}_i \frac{1}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} \tag{67}$$

The quantity $\mathcal{Z}_c$ is the *characteristic wave impedance* of the $TE_{1,0}$ wave. For air or space as the dielectric medium

$$\mathcal{Z}_c = 377 \frac{\lambda}{\lambda_0} = 377 \frac{1}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} \quad \text{ohms} \tag{68}$$

The characteristic wave impedance of Eqs. 67 and 68 corresponds to the characteristic wave impedance $\mathcal{Z}_c = \mathcal{Z}_i$ of the parallel-strip line (see Art. 2) and, less closely, to the characteristic resistance R_c of the line. As pointed out in Art. 2, $\mathcal{Z}_c$ must be distinguished from R_c—$TE_{1,0}$ waves in rectangular guides of the same a dimensions but unequal b dimensions have the same characteristic wave impedances; however, a matched condition is not obtained if two such guides are joined. Furthermore, the characteristic resistance of a transmission line is independent of frequency, whereas the value of $\mathcal{Z}_c$ is altered by a change of frequency. Observe that in the definition of characteristic wave impedance only the field components in the *transverse plane* are employed. As explained in Art. 7, only these components are associated with the traveling-wave aspect of waveguide fields.

Sometimes it is helpful to associate with a waveguide a characteristic impedance that is a ratio of voltage to current. Such impedances may

be defined in a number of ways,[1] but their usefulness is limited because their definition requires rather arbitrary choices of voltage and current. Only the wave-impedance concept is employed in this book.

If the standing-wave ratio for the guide in Fig. 25*A* is not unity, the electric- and magnetic-field intensities at any cross section must be obtained by adding the field intensities of the incident and reflected waves. The variations of the resultant intensities $\mathcal{E}$ and $\mathcal{H}_x$ are, in general, no longer in phase. The relation of $\mathcal{E}$ to $\mathcal{H}_x$ within any cross section may be described by a complex-number wave impedance $\mathcal{Z}$. The magnitude of $\mathcal{Z}$ indicates the peak values of $\mathcal{E}$ divided by that of $\mathcal{H}_x$, and the angle of $\mathcal{Z}$ denotes the phase of $\mathcal{E}$ relative to that of $\mathcal{H}_x$. The quantity $\mathcal{Z}$ is the *input wave impedance* for the section of the guide to the right (in Fig. 25*A*) of the cross section for which it specifies the $\mathcal{E}$ to $\mathcal{H}_x$ relation.

The relation of incident and reflected waves to the resultant standing wave is exactly the same for the field intensities in a guide as it is for the voltage and current of a transmission line. Therefore, a complex-number reflection coefficient Γ may be assigned to each cross section of a waveguide to express the magnitude ratio and phase difference for incident and reflected electric-field-intensity waves at that cross section. The relation of Γ to $\mathcal{Z}/\mathcal{Z}_c$ is the same as the relation of the transmission-line reflection coefficient to $Z/R_c = z$, the per-unit impedance. Thus, if

$$z = \frac{\mathcal{Z}}{\mathcal{Z}_c} \tag{69}$$

is defined as the per-unit impedance at any cross section of the guide, the corresponding reflection coefficient is given (see Eqs. 37 and 39, pp. 495 and 496. of Chap. VII) by

$$\Gamma = \frac{z - 1}{z + 1}$$

The Smith chart, p. 500, is the graphical expression of this relation and can therefore be used to relate reflection coefficient to per-unit impedance in waveguides.

The Smith chart provides a convenient means of determining the per-unit impedance at any cross section of a guide from standing-wave measurements. The procedure is exactly the same as that used with transmission-line waves. At a point of minimum electric-field intensity (or maximum magnetic-field intensity), the per-unit impedance is resistive and equal to $1/\rho$, where ρ is the standing-wave ratio. A point on the Smith chart is thus determined. By an appropriate rotation of this point, the per-unit impedance at any cross section may be found. The wave impedance at the cross section may be calculated from the cor-

[1] S. A. SCHELKUNOFF, *Electromagnetic Waves*, 319 (D. Van Nostrand Company, Inc., New York, 1943).

responding per-unit impedance and the characteristic wave impedance of the guide.

Standing-wave measurements of fields in a region free of higher order waves may be used in this manner to determine the per-unit impedance z_L of the load device (see Fig. 25*A*) and the per-unit input impedance z_d of the entire guide, which quantity is the per-unit impedance that the source faces. These quantities are useful because the first expresses the effect of the load upon the waveguide and the second indicates the effect of the guide upon the source—the power delivered by the source, for example, is a function of the per-unit input impedance.

The wave impedance of a load and the wave impedance faced by a source can be calculated from the corresponding per-unit impedances. The wave impedances, however, are of little use and are seldom determined. The wave impedance specifies the ratio of field intensities for the $TE_{1,0}$ wave and has no direct physical significance in the load and source regions, where the higher order waves are important. Moreover, the wave impedance of a load, for example (in contrast to the load impedances of circuit theory), is not a property of the load alone but applies to a *load used with a particular size and shape of guide.* The wave impedance is thus no more characteristic of the load than is the per-unit quantity, and there is no advantage in extending calculations beyond the determination of the per-unit impedance.

Typical load and source devices have dimensions large relative to a guide wavelength. Thus the distances required in calculating load and input per-unit impedances from standing-wave measurements become difficult to determine. To permit accurate specification of the system, *reference planes* must be assigned to the load and the source, as in Fig. 25*A*, and per-unit impedances calculated with reference to these planes. The positions of the reference planes may be chosen arbitrarily, and the impedance values obtained depend upon the location of these planes. To specify the effect of a load on a waveguide, the per-unit impedance and the reference-plane location must be stated.

As an example of waveguide impedance calculations, assume the guide in Fig. 25*A* to have outside dimensions of 3 by 1½ in. and an a dimension (see Table 1) of 7.22 cm. The guide is 120 cm long between reference planes and is operated at 3,000 Mcps ($\lambda_0 = 10$ cm). Standing-wave measurements in the central region indicate a standing-wave ratio $\rho = 3$ and a position of minimum electric-field intensity 40 cm from the load reference plane. The cutoff wavelength $\lambda_{co} = 2a$ is 14.44 cm, and the guide wavelength, by Eq. 41, is

$$\lambda = \frac{\lambda_0}{\sqrt{1 - (\lambda_0/\lambda_{co})^2}} = \frac{10}{\sqrt{1 - (10/14.44)^2}} = 13.9 \text{ cm}$$

Thus the distance from the load to the measured point of minimum electric-field intensity is $40/13.9 = 2\frac{7}{8}$ wavelengths, and from the minimum-electric-field point to the source is $80/13.9 = 5\frac{3}{4}$ wavelengths.

The per-unit values of input impedance and load impedance may be found from these data. The source reference plane, located an odd number of quarter wavelengths from the electric-field minimum, is at a position of maximum electric-field intensity. The per-unit input impedance is thus resistive and equal to ρ, which has the value 3. This impedance may also be found by turning clockwise from point A at $0.333 + j0$ in the Smith chart in Fig. 26 through an angle corresponding to $5\frac{3}{4}$ (or $\frac{1}{4}$) wavelengths to point B at $3 + j0$. Turning counterclockwise from point A through an angle corresponding to $2\frac{7}{8}$ (or $\frac{3}{8}$) wavelengths yields at point C the per-unit load impedance which is $0.6 + j0.80$.

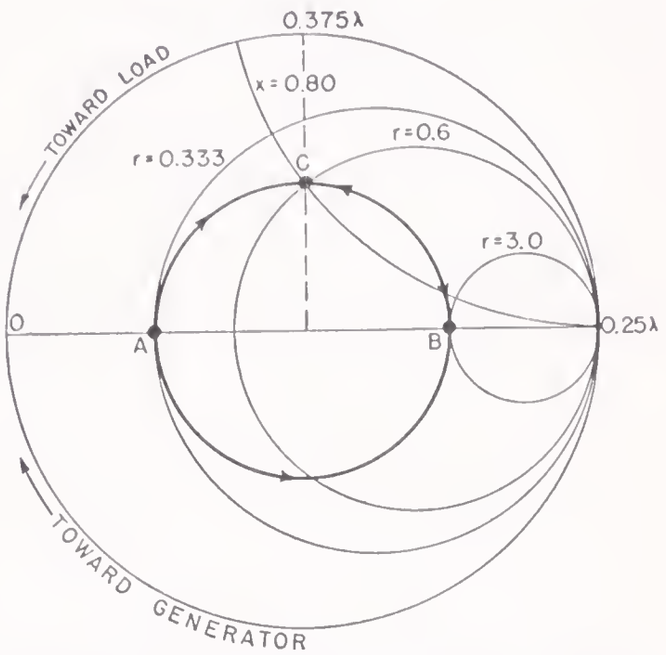


FIG. 26. Construction on Smith chart for example based on Fig. 25.

The characteristic wave impedance of the guide Z_c is $377 \times 13.9/10$, or 525 ohms. The wave impedance varies from one cross section to another from a value of $525 \times 3 = 1{,}575$ ohms, resistive, at a point of maximum electric-field intensity, through inductive and capacitive values to $525/3 = 175$ ohms, resistive, at a point of minimum electric-field intensity.

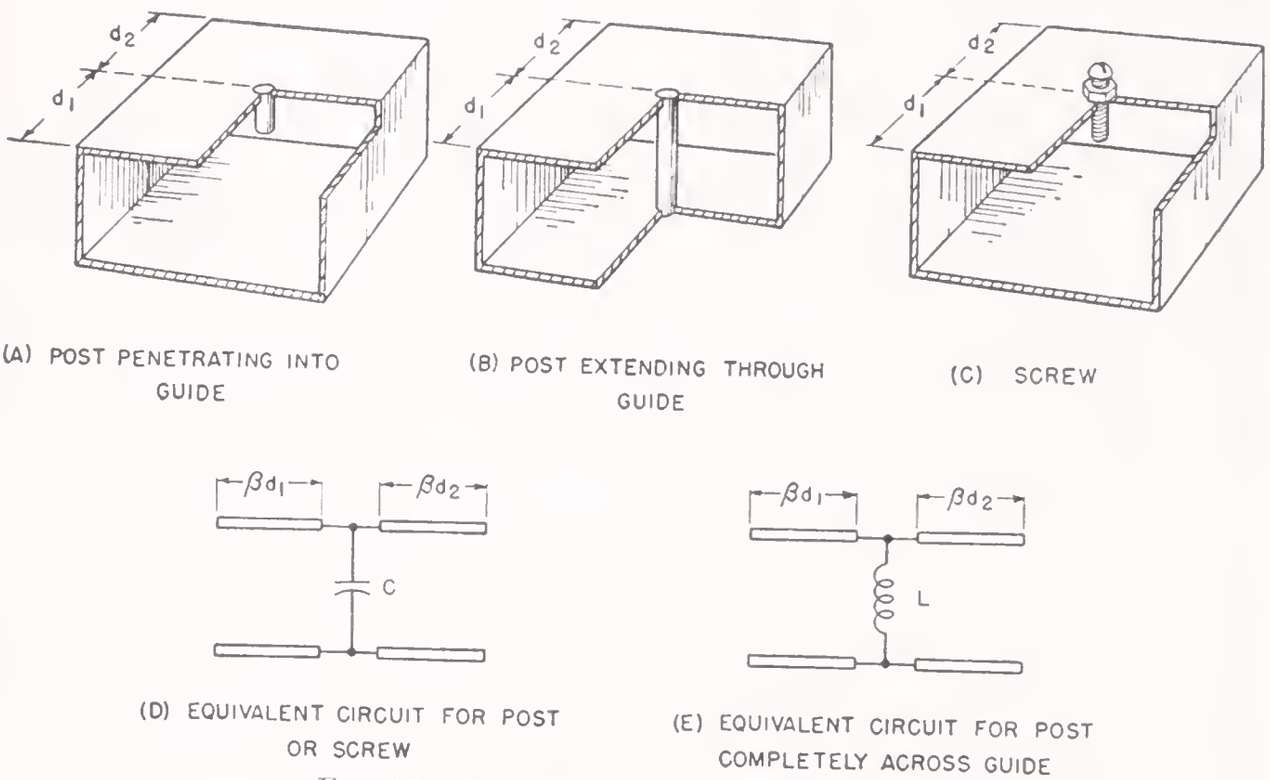
14. Impedance-changing Devices. Often in radar, r-f transmission systems are not perfectly matched by the load devices. The standing waves that result are undesirable in waveguides, as they are in two-wire lines, because of the associated increase in power loss, reduction of power capacity, and increase of frequency sensitivity. Impedance-changing devices are therefore introduced near the sources of reflected waves to provide a standing-wave ratio nearly equal to unity over the major part of the system. Impedance-changing devices are employed also near the source to provide a load impedance for the source of optimum value.

A waveguide device analogous to the short-circuited stub used in impedance transformers for two-conductor lines would be a length of guide closed at one end by a conducting wall. Such waveguide stubs are not employed to control impedances in waveguide systems because simpler devices are available. The simpler devices are *posts* or *screws* and *irises*. These devices perform the same function as do the transmission-line

stubs: They provide shunt susceptances at appropriate points in the equivalent circuit of the waveguide system.

Waveguide posts are conducting rods that enter the dielectric space in a guide, as in Fig. 27*A*, or extend completely across the guide, as in Fig. 27*B*. A post of adjustable depth of penetration can be constructed in the form of a matching screw, as in Fig. 27*C*.

The effect of a post on the waveguide may be explained in one way if the guide (in which only $TE_{1,0}$ waves are propagated) is thought of as similar to a parallel-strip line, except that the voltage between conductors and the longitudinal current vary sinusoidally from a maximum at the



(A) POST PENETRATING INTO GUIDE

(B) POST EXTENDING THROUGH GUIDE

(C) SCREW

(D) EQUIVALENT CIRCUIT FOR POST OR SCREW

(E) EQUIVALENT CIRCUIT FOR POST COMPLETELY ACROSS GUIDE

FIG. 27. Posts and screws in waveguides.

center to zero at the edges. The short post of Fig. 27*A* or the screw of Fig. 27*C* may then be likened to a capacitor plate extending from one transmission-line conductor. In this way, the equivalent circuit of Fig. 27*D* is suggested. Similarly, the post of Fig. 27*B* resembles a conducting wire joining the line conductors. The post is not a perfect short circuit because it is encircled by a magnetic field and possesses an inductive reactance, as indicated by the equivalent circuit of Fig. 27*E*. The *susceptances* of the equivalent capacitor and inductor in Figs. 27*D* and *E* decrease if the diameter of the post is decreased or if the post is moved from the center toward one side of the guide into a region of weaker electric fields.

These conclusions perhaps require interpretation in terms of the meaning of the equivalent circuit. As explained in Art. 13, the transmission-line symbol used in a waveguide equivalent circuit indicates only that the dominant-wave fields in the guide are described by a standing-wave

pattern obtained by summing an incident and a reflected wave. The equivalent circuits of Figs. 27*D* and *E*, therefore, indicate that two different standing-wave patterns, both comprising dominant waves only, must be considered, one on either side of the post. Because the post is represented as a *shunt* element, the electric-field intensity (corresponding to voltage) at the post is the same in the two patterns, whereas the magnetic-field intensity (corresponding to current) of the patterns makes an abrupt change at the post.

Justifying the shunt connection directly from the field distribution is difficult because fields near the post are the sum of dominant waves and higher order waves generated at the post, and only the dominant waves are accounted for in the equivalent circuit. It may be seen, however, that longitudinal components of guide-wall currents approaching the post from the source must account for the sum of the current in the post and the longitudinal current leaving the post. Thus the post is associated with a change in longitudinal currents and a corresponding change in the transverse magnetic-field intensity.

The nature of the shunt element in the equivalent circuit may be determined by considering the fields of the higher order waves near the posts. These fields represent the difference between the total field intensities and the sum of the field intensities of the dominant waves. For the short posts of Figs. 27*A* and *C*, the additional field is mainly an electric field, with field-intensity lines leaving the end of the post in all directions. Corresponding to the electric field, a capacitance appears in Fig. 27*D*. Similarly, the post of Fig. 27*B* is encircled by a magnetic field, and the equivalent circuit contains an inductance. Alternatively, for Fig. 27*A* it may be observed that the electric-field intensity has a longitudinal component near the post. Thus higher order transverse-magnetic waves are generated, and the net stored energy is electric-field energy (see Art. 12). In Fig. 27*B*, the magnetic-field intensity of the higher order waves has a longitudinal component; therefore, higher order transverse-electric waves are generated, and the stored energy is predominantly magnetic.

The higher order waves in Fig. 27*A* are not entirely transverse-magnetic because the conduction current in the post and the displacement current below the post result in magnetic-field lines encircling the post and the electric-field lines. Because of the inductive-reactance voltage generated by the rate of change of the magnetic flux, a more detailed equivalent circuit, which is applicable for any depth of penetration of the post, is the circuit of Fig. 28*A*.

For small penetration depths, the reactance of L is small and that of C is large, and the net reactance is capacitive. With increase of penetration, inductance and capacitance both increase, and at a critical penetration series resonance occurs. The guide is then effectively short-cir-

cuited; beyond the post only the higher order waves are present, and these attenuate rapidly with distance. Penetration depths greater than the resonant value cause the inductive reactance to predominate, and the capacitance of Fig. 28*A* becomes a short circuit when the post extends from wall to wall. Typical variations of reactance and susceptance associated with the change of penetration depth are indicated in Fig. 28*B*. The penetration distance for series resonance is approximately $\lambda_0/4$.

The reactance of a matching device depends upon the guide dimensions as well as upon the device itself and therefore, like the impedance of a load

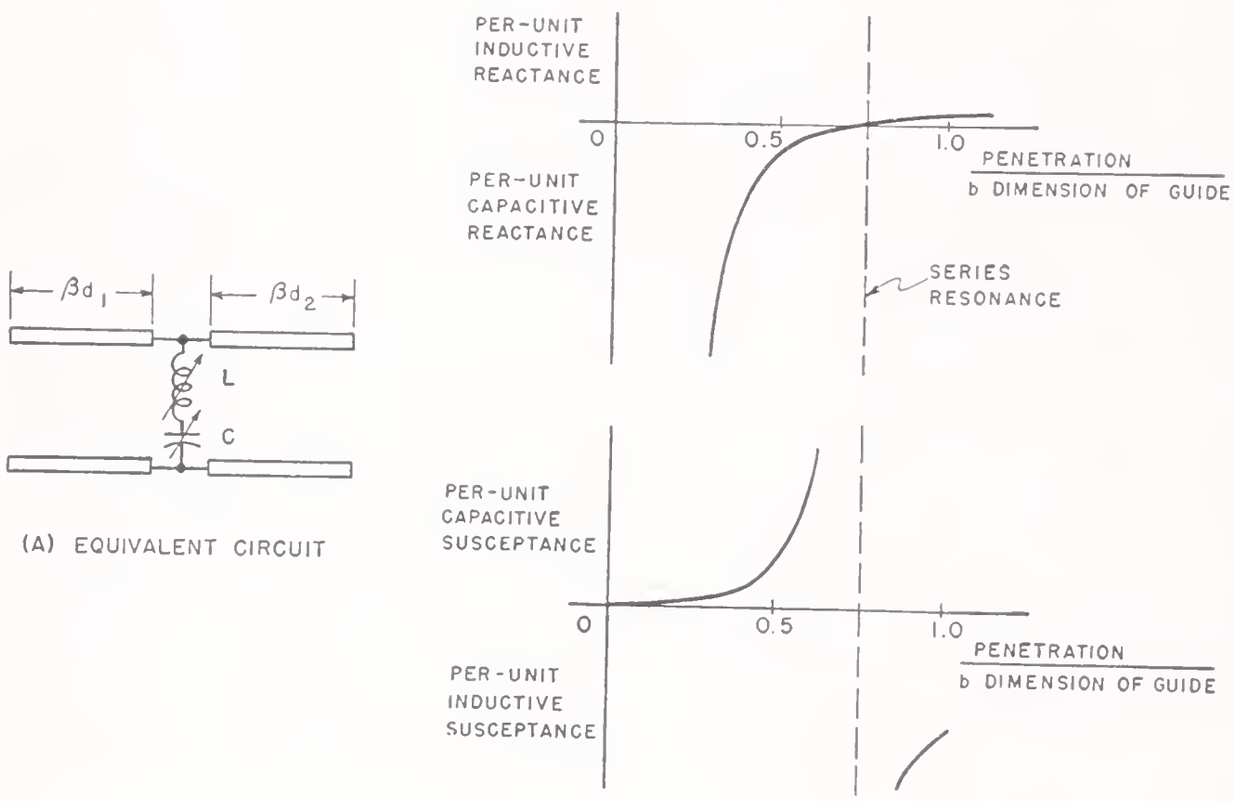


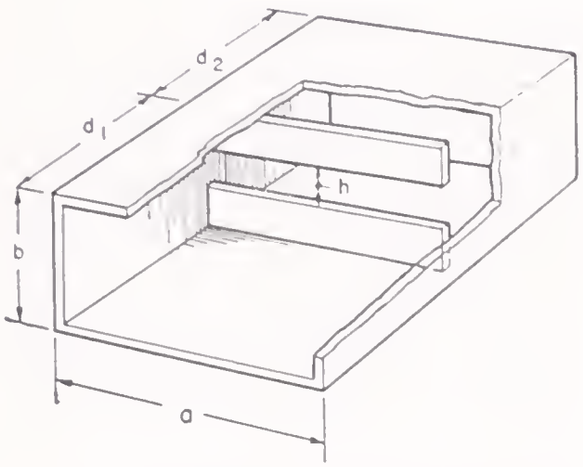
FIG. 28. Effect of penetration of post.

device, is always expressed as a per-unit quantity. Furthermore, the value of the per-unit reactance (and, in fact, the structure of the equivalent circuit) is dependent upon the choice of reference planes. For all the equivalent circuits of Figs. 27 and 28, a reference plane is taken through the center of the post, as indicated by the dimensions d_1 and d_2 in Fig. 27. The transmission-line symbols in the equivalent circuit represent guide sections on either side of this post extending to the reference plane.

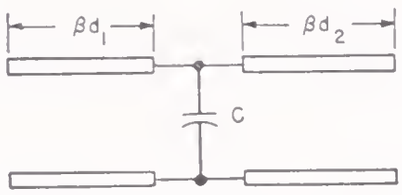
An iris is a metal diaphragm located in a transverse plane of a waveguide and containing openings (called *windows* or *apertures*) through which the waves may pass. If the aperture divides the diaphragm into two parts, one attached to the top of the guide and one to the bottom, as in Fig. 29*A*, a *capacitive iris* is formed. If the window divides the

diaphragm into parts attached to the side walls, as in Fig. 30*A*, an *inductive iris* is produced.

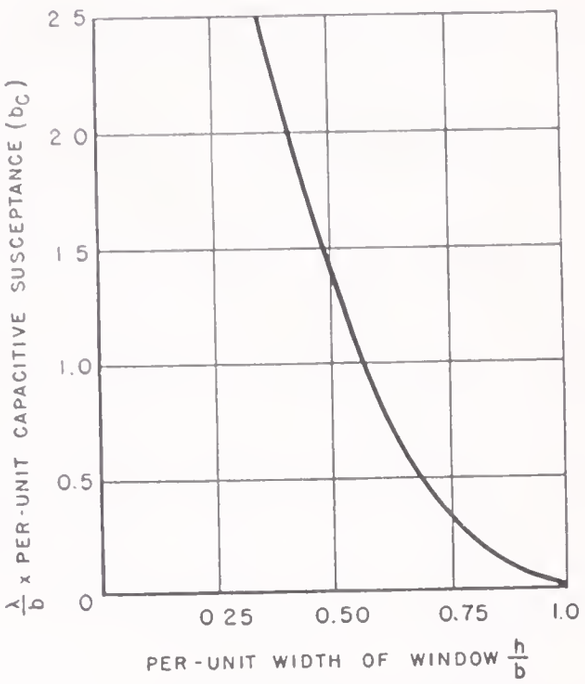
The equivalent circuit for the capacitive iris (Fig. 29*B*), which is drawn for a reference plane through the center of the iris, may be explained in much the same manner as that of the short matching post. The field added to the dominant-wave fields in the vicinity of the iris comprises mainly the lines of electric-field intensity extending between edges of the iris and can be represented by transverse-magnetic higher order waves. Thus the reactive element in the equivalent circuit is a capacitance. An



(A) PICTORIAL REPRESENTATION



(B) EQUIVALENT CIRCUIT



(C) SUSCEPTANCE OF THIN IRIS

FIG. 29. Symmetrical capacitive iris.

important difference between this iris and the matching post is that no higher order transverse-electric waves are generated by the iris. The net stored energy is always electric-field energy, and decreasing the window height (h in Fig. 29*A*) causes a continual increase of capacitive susceptance, until for $h = 0$ the guide is short-circuited by being completely blocked by the iris.

Calculation of the susceptance of the capacitor in the equivalent circuit of Fig. 29*B* is possible, but the procedure is somewhat lengthy.[1] For an iris of negligible thickness having an opening of height h located symmetrically between top and bottom of a guide of height b, the per-unit

[1] J. W. MILES, "The Equivalent Circuit for a Plane Discontinuity in a Cylindrical Wave Guide," *Proc. I.R.E.*, **34** (October, 1946), 728–742.

capacitive susceptance b_c depends upon the ratio h/b and the ratio b/λ in the following manner:

$$b_c = 4\frac{b}{\lambda}\ln \csc \frac{\pi}{2}\frac{h}{b} \tag{70}$$

The product of b_c by λ/b therefore depends only on h/b, and its value varies as indicated in Fig. 29*C*. If the iris is of appreciable thickness, the susceptance is somewhat increased.

If apertures are cut in an iris in such a way that one or more conductors extend from top to bottom of the guide, magnetic fields similar to those

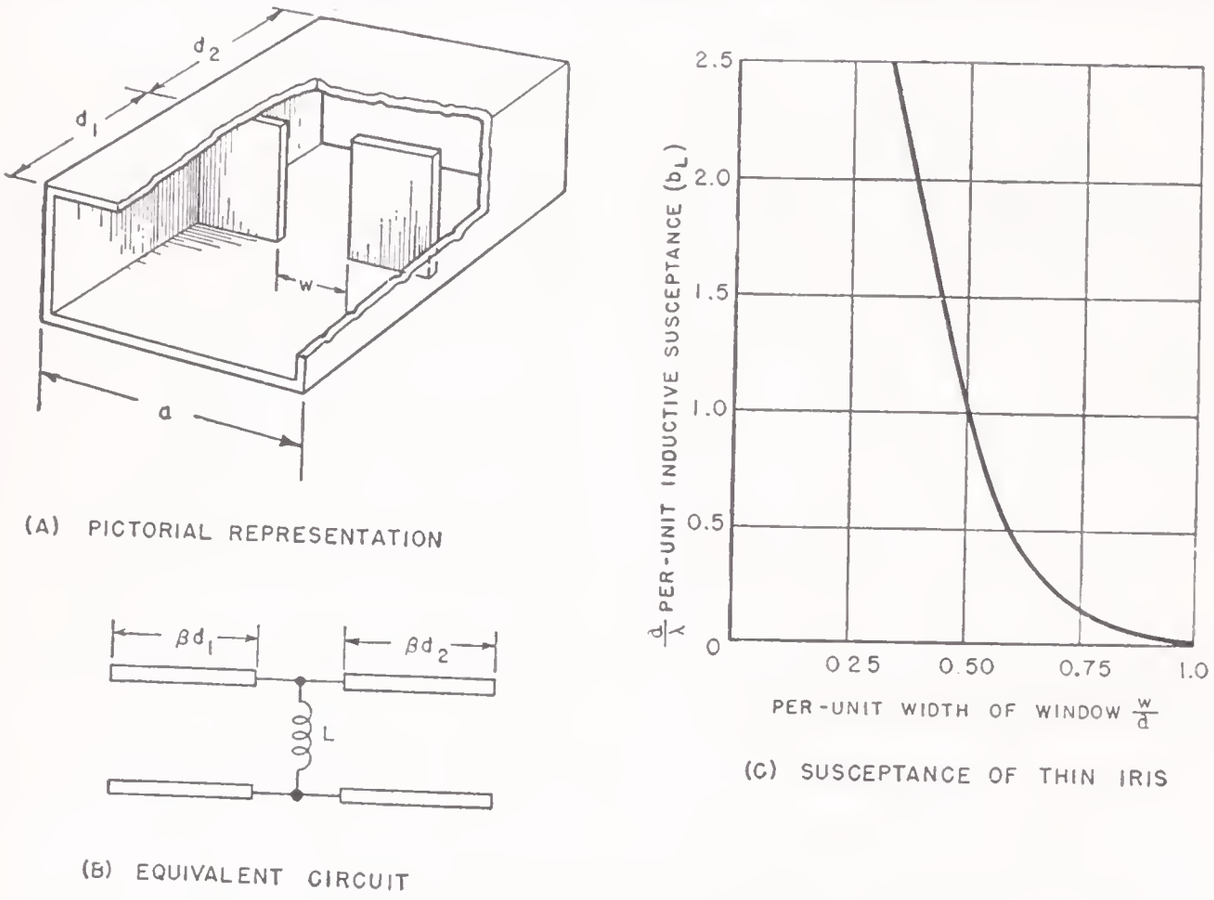


FIG. 30. Symmetrical inductive iris.

associated with the post of Fig. 27*B* surround the conductors. Higher order waves of the transverse-electric type are generated, and the iris is inductive. A particular form of inductive iris in which a single window is located symmetrically between the side walls is illustrated in Fig. 30*A*, and an equivalent circuit for a reference plane at the center of the iris appears in Fig. 30*B*. The per-unit inductance susceptance b_L of a very thin iris of this form depends upon the ratio of w (the width of the window) to the a dimension of the guide and upon the ratio λ/a, in accordance with the relation

$$b_L = \frac{\lambda}{a}\cot^2 \frac{\pi}{2}\frac{w}{a} \tag{71}$$

The product $b_L(a/\lambda)$ depends only upon the ratio w/a, and the nature of the dependence is shown by Fig. 30*C*. The per-unit susceptance is slightly increased if the thickness of the iris is appreciable.

The inductive iris and the post of Fig. 27*B* have an advantage over the capacitive devices in high-power waveguide systems. The capacitive devices produce a region of intensified electric field and thus reduce the power capacity of the system. In some waveguide devices, *resonant irises* combining the effects of inductive and capacitive irises into one unit are employed. For example, for an iris containing a rectangular window of width less than the a dimension and height less than the b dimension, the equivalent circuit is a parallel combination of inductance and capacitance having a resonant frequency controllable by adjustment of the window dimensions.

The manner in which posts or irises may be used for impedance matching in a waveguide may be illustrated by an example. Consider again the waveguide system of Fig. 25*A*, p. 596, operating under the conditions specified at the end of Art. 13. An inductive iris is to be added to this system in order to provide unity standing-wave ratio over the major portion of the guide.

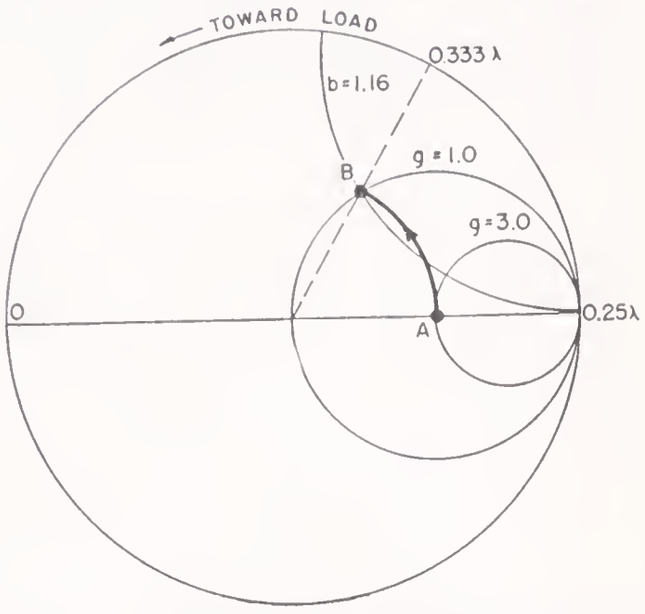


FIG. 31. Construction on Smith chart for example illustrating use of inductive iris for elimination of standing waves.

At the minimum-electric-field position 40 cm from the load, the per-unit input susceptance of the portion of the guide leading to the load is a conductance numerically equal to the standing-wave ratio. This conductance is represented by point A at $3 + j0$ on the Smith chart of Fig. 31. The input susceptance at points closer to the load is found by moving counterclockwise along a $\rho = 3$ circle, and at point B the conductance component is unity. Thus a susceptance at a position corresponding to B can be used to cancel the imaginary portion of the admittance at B and yield unity standing-wave ratio between the source and this point. The admittance at B is $1 + j1.16$; thus an inductive iris having a per-unit admittance $-j1.16$ is needed to make the total admittance $1 + j0$.

If a very thin symmetrical iris is employed, the aperture width w of Fig. 30 can be determined from Eq. 71. For the values of this example, this equation becomes

$$1.16 = \frac{13.9}{7.22} \cot^2 \frac{\pi}{2} \frac{w}{7.22}$$

and thus the per-unit width is $w/a = 0.58$, and the actual width is

4.18 cm. The proper location of the iris is determined from the angular distance between points A and B in Fig. 31, which corresponds to $0.333 - 0.25 = 0.083$ wavelength or 1.15 cm. The iris may therefore be located $40 - 1.15 = 38.85$ cm from the load reference plane. Of course, other locations at half-wave intervals provide the same impedance relations, and locations closer to the load restrict the standing waves to a shorter portion of the waveguide. For maximum length of flat waveguide, therefore, the position $38.85 - 5 \times 13.9/2 = 4.10$ cm from the load reference plane should be used. Observe, however, that, for a matching device so close to the load, the regions of higher order waves near the load and near the matching device may be expected to overlap. Under such circumstances, the higher order waves interact, and the resulting coupling effect modifies the per-unit admittance of iris and load.

15. Circular Waveguides. Although rectangular waveguides are used almost exclusively in radar systems, guides of other cross-sectional shapes

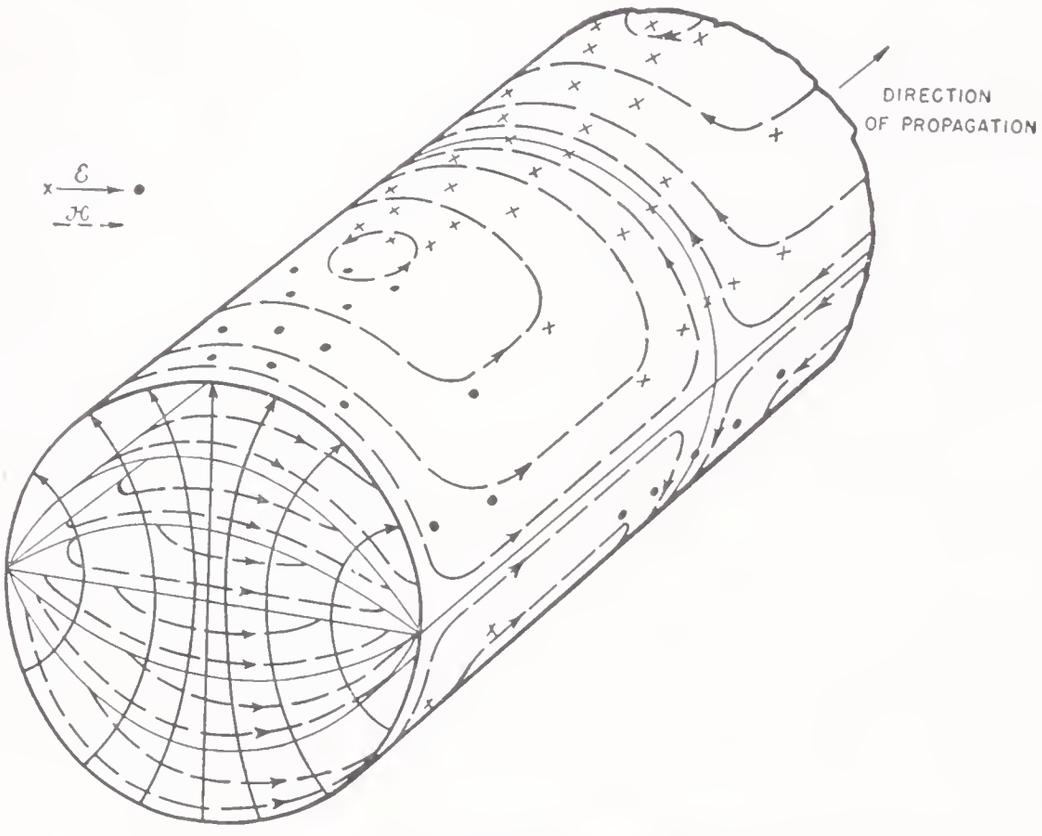


FIG. 32. The $TE_{1,1}$ wave of a circular waveguide.

are equally capable of transmitting radio-frequency power. For each cross section there exists a variety of wave types that may be transmitted. The wave of lowest cutoff frequency is called the dominant wave, and the others are the higher order waves. Guides of circular cross section are useful in special applications in radar r-f systems.

The dominant wave of a circular guide is a transverse-electric wave designated $TE_{1,1}$ and having the pattern illustrated in Fig. 32. Observe that this is the pattern obtained if the field configuration of the dominant

rectangular wave is distorted in such a way that the guide cross section becomes circular. The cutoff wavelength of the round guide is 1.71 times the diameter of the guide. Thus the diameter must be $2/1.71 = 1.17$ times the a dimension of a rectangular guide having the same cutoff frequency. The circular guide therefore has the disadvantage of somewhat greater bulk and weight, accompanied by advantages of greater power capacity and lower attenuation because of the greater size. Perhaps the most important reason for the general preference for rectangular guides is that the rectangular cross section definitely fixes the polarization of the wave transmitted. In the circular guide, minor irregularities of the wall

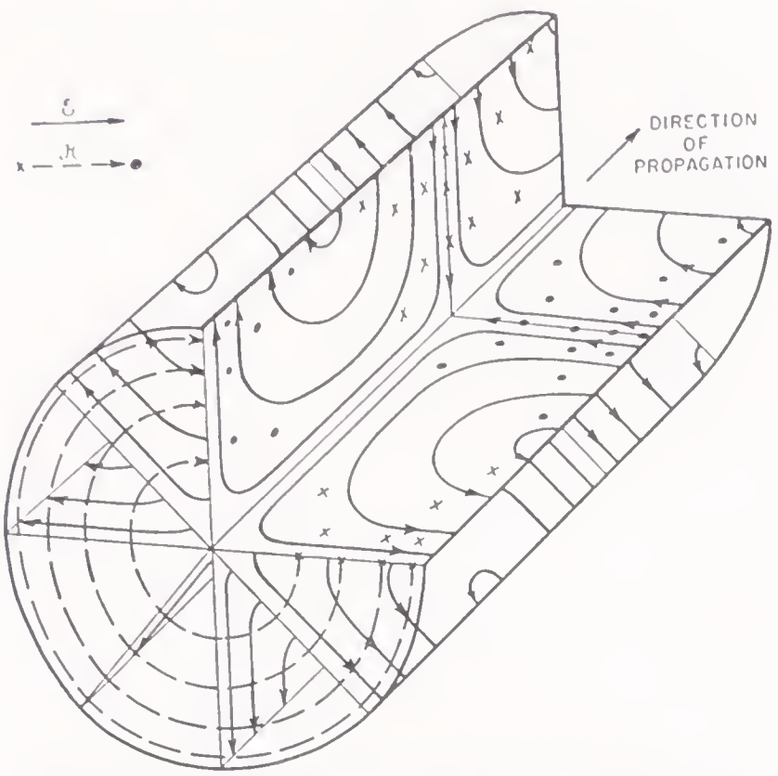


FIG. 33. The $TM_{0,1}$ wave of a circular waveguide.

surface might cause the field pattern to turn within the guide, thus disturbing the operation of the terminating devices.

Certain higher order waves of circular guides have special properties of importance. Among these waves is the type $TM_{0,1}$ (Fig. 33), which has the longest cutoff wavelength—1.31 times the guide diameter—of any wave except the dominant wave. As shown by Fig. 33, the field pattern has circular symmetry. Thus, if a terminating device designed to radiate or receive this wave is rotated about the guide axis, no change in the electrical characteristics of the device occurs. A short length of circular guide carrying this wave is therefore useful if one part of a waveguide system is required to rotate relative to another part—as is necessary, for example, if a radar antenna is to turn while transmitter and receiver remain stationary. Rotary joints of this and other forms are described in Art. 3, Chap. XII.

The $TM_{0,1}$ wave is one of a group of *circular waves*. In the circular

transverse-magnetic ($TM_{0,m}$) wave, the magnetic-field lines form circles in the transverse plane, as illustrated by Fig. 33. In circular transverse-electric ($TE_{0,m}$) waves, the electric-field lines form similar circles, as illustrated by the $TE_{0,1}$ and $TE_{0,2}$ patterns in Fig. 34. The $TE_{0,m}$ waves

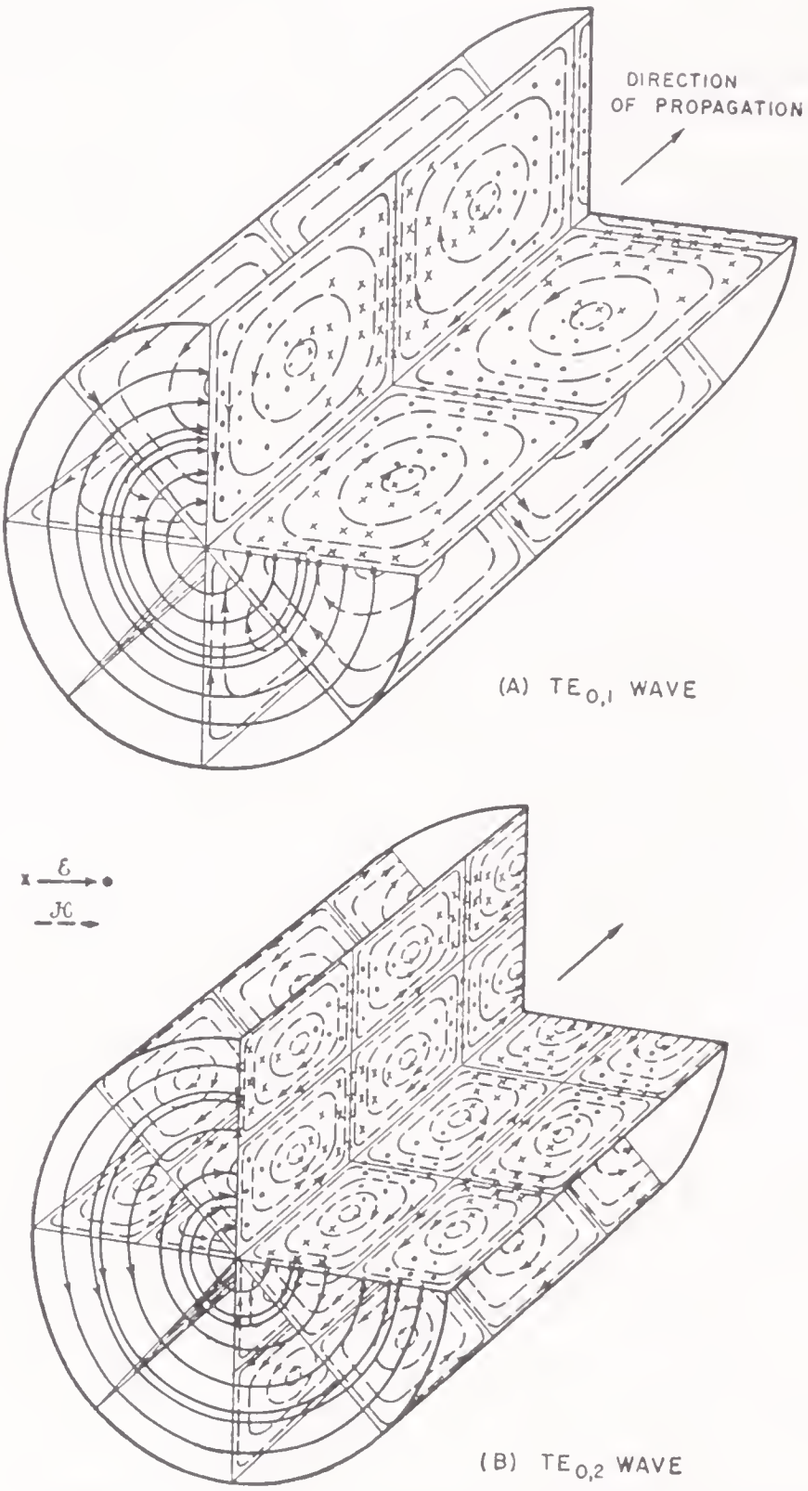


FIG. 34. Circular transverse-electric waves of a circular waveguide.

are remarkable in that none of the electric-field lines terminate at the walls of the guide. A related condition is that the magnetic field at the wall surfaces is entirely longitudinal, and therefore only transverse currents flow in the guide walls. In Art. 9, the attenuation of the rectangular-guide $TE_{1,0}$ wave was found to result from both transverse and longitudinal components of currents in the guide walls. It may be sur-

mised, therefore, that elimination of the longitudinal currents by use of a $TE_{0,m}$ wave should decrease the attenuation caused by imperfectly conducting walls.

The attenuation of the $TE_{0,m}$ waves is, in fact, extremely small and has the property of decreasing indefinitely as the frequency of the source is increased relative to the cutoff frequency. (Contrast the attenuation curve for the dominant rectangular-guide wave, plotted in Fig. 21, p. 587.) The reason for the decrease of attenuation with frequency is that in any transverse-electric wave the ratio of longitudinal to transverse magnetic field decreases with increase of frequency. (See, for example, Eqs. 44 and 45, p. 575, which apply to the dominant rectangular wave.) Thus, for constant transmitted power, increasing the frequency decreases the magnetic field at the side walls and causes a corresponding decrease in conduction currents and attendant dissipation.

These low-attenuation circular waves are of interest in communications because they provide a possibility of carrying signals great distances with small decrease in amplitude. In radar, short sections of guide terminated at both ends by conducting plates are useful as resonant elements (see Art. 5, Chap. IX), and extremely high values of Q are possible if low-attenuation waves are employed in the resonators. Although low-attenuation waves provide a pattern of circular symmetry, just as does the $TM_{0,1}$ wave, they are never employed in rotary joints. In the short length of guide required by the joint, the low attenuation is unimportant, and the design of terminating devices is much more difficult for any $TE_{0,m}$ wave than for the $TM_{0,1}$ wave. The $TE_{0,1}$ wave has a relatively short cutoff length, 0.82 times the guide diameter, and four waves of equal or greater length may be propagated simultaneously unless the terminating devices are carefully designed. When the $TM_{0,1}$ wave is employed, only the dominant wave can interfere.

The method of designating wave types in circular guides is, in a sense, similar to that used for waves in rectangular guides. As explained in Art. 10, if neither m nor n is zero, the patterns of $TE_{m,n}$ or $TM_{m,n}$ waves may be considered to be formed by stacking a number of identical small rectangular guides into a close rectangle m guides wide and n guides high. Each of the small guides carries a $TE_{1,1}$ or $TM_{1,1}$ wave, and removal of the interior walls of the array yields the desired pattern.

In a corresponding manner, a circular waveguide may be split into sections by planes through the axis of the guide and circular cylinders coaxial with the guide, as illustrated by the cross section of Fig. 35*A*. Field patterns that are a distortion of the $TE_{1,1}$ or $TM_{1,1}$ wave of the rectangular guide are then assigned to the small curved guides produced (except that for transverse-electric waves a distortion of the rectangular $TE_{1,0}$ pattern is used in the central sectors). The transverse-electric and

transverse-magnetic fields thus built up are represented in Figs. 35*B* and *C*. Observe that electric-field lines are perpendicular to all boundary planes and magnetic-field lines are parallel to the boundary planes. The resulting wave is designated $TE_{n,m}$ or $TM_{n,m}$, where n is the number of axial planes inserted to divide the guide and m the number of cylinders inserted, plus one for the guide itself.

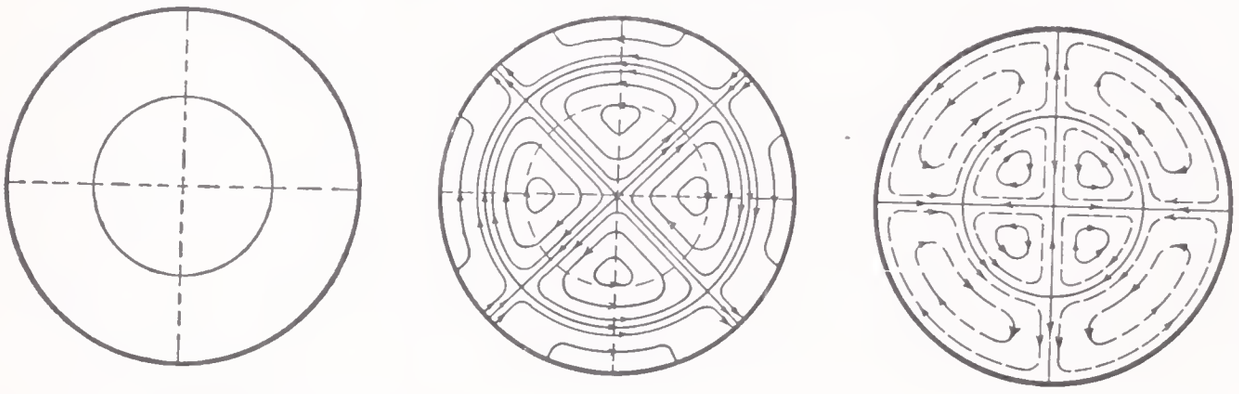


FIG. 35. Diagrams illustrating method of designating wave in circular waveguides.

Of the waves in Figs. 32 to 34, only the $TE_{1,1}$ wave is included in the above classification. For this wave (Fig. 32), the number of cylinders is unity (the guide itself), and the number of planes that can be passed through the axis perpendicular to all the electric-field lines is also unity. The waves in Figs. 33 and 34 are circular waves and may be considered the forms that result when no dividing planes are used. Thus these forms are designated $TE_{0,m}$ or $TM_{0,m}$, where m is, as before, the number of conducting cylinders (plus one for the guide) that can be inserted to divide the pattern.

CHAPTER IX

RESONANT LINES AND CAVITIES

Circuits composed of lumped inductance and capacitance elements can be made to resonate at any frequency from less than one cycle per second to many hundreds of megacycles per second. At extremely high frequencies, however, the physical size of the inductors and capacitors becomes so small and the losses in the circuit become so great that resonant devices of different construction are preferred. In the uhf range, sections of parallel-wire or coaxial transmission line are commonly employed in place of lumped-constant resonant circuits, and in the microwave region, cavity resonators are used. Cavity resonators are hollow metal-walled chambers fitted with devices for admitting and extracting electromagnetic energy.

Although resonators built for different frequency ranges have a variety of physical forms, the basic principles of operation are essentially the same for all. Operating principles of resonators are explained in this chapter, and these principles are applied in subsequent chapters to the study of important radar-system components employing resonators. These components include transmitting oscillators, receiver local oscillators and input circuits, transmit-receive devices, and various auxiliary devices used in r-f transmission systems.

1. Resonant Modes of Networks. All resonators, whether reactive networks, transmission lines, or cavities, have certain *natural frequencies* of oscillation. A natural frequency is a frequency for which oscillations in a lossless resonator can be sustained, once started, even though the resonator is completely isolated from all sources of power. Some resonators have only a single natural frequency; others may have several or even infinitely many. Each natural-frequency oscillation is called a *natural mode* or *resonant mode*. The term *mode* signifies a *manner of oscillation*. The various modes occur at different frequencies and in general cause different patterns of voltage and current or of electric and magnetic fields in the resonator. A description of a natural-mode oscillation should therefore include a specification of (1) the frequency, (2) the voltage and current patterns, or the electric- and magnetic-field patterns, and (3) a measure of the amplitude of the oscillations.

Although this chapter pertains primarily to high-frequency resonators, it is desirable to consider the high-frequency devices in relation to lumped-

constant L-C networks operating as resonators. Figure 1 shows the simplest network resonator—a tuned circuit consisting of a single inductor and capacitor. In this circuit, an oscillation once started (by charging the capacitor before connecting it to the inductor, for example) will persist indefinitely if the elements are lossless. The frequency must be such that L and C carry equal currents and have equal terminal voltages; that is, the inductive and capacitive reactances must be equal. Thus $\omega_0 L = 1/\omega_0 C$, where $\omega_0 = 2\pi f_0$ is the angular frequency of the oscillation. The oscillation frequency f_0 is therefore

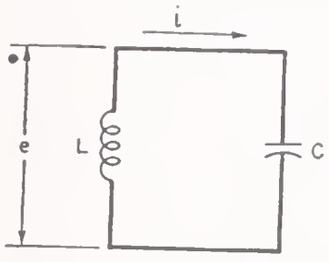


FIG. 1. Resonator composed of one inductor and one capacitor.

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{1}$$

Because the reactances are equal at only a single frequency, the circuit of Fig. 1 has but one natural frequency and hence one mode of oscillation.

The oscillations in the network of Fig. 1 involve only a single voltage and a single current, denoted by e and i. Both voltage and current vary sinusoidally with time, and the mode pattern may therefore be described by specifying the phase and magnitude of e and i during the oscillation. For the reference directions shown, e leads i by 90 deg, and E_m, the peak value of the voltage, equals I_m, the peak value of the current, multiplied by the reactance of either the capacitor or inductor; that is,

$$\frac{E_m}{I_m} = \frac{1}{\omega_0 C} = \omega_0 L = \sqrt{\frac{L}{C}} \tag{2}$$

The energy stored in the network at any instant is the sum of the energy stored in the inductance and the energy stored in the capacitance. That is, the total stored energy W is

$$W = \tfrac{1}{2}Li^2 + \tfrac{1}{2}Ce^2 = \tfrac{1}{2}LI_m^2 \cos^2 \omega_0 t + \tfrac{1}{2}CE_m^2 \sin^2 \omega_0 t$$

Because $E_m^2 = I_m^2 L/C$,

$$W = \tfrac{1}{2}LI_m^2 \cos^2 \omega_0 t + \tfrac{1}{2}LI_m^2 \sin^2 \omega_0 t = \tfrac{1}{2}LI_m^2 \tag{3}$$

or, in terms of E_m and C,

$$W = \tfrac{1}{2}CE_m^2 \tag{4}$$

Thus the total energy stored in the circuit is *constant* and is equal to the peak energy in either L or C.

That the energy in the system should not change during the cycle of oscillation is to be expected, because the system is assumed to be lossless and isolated from any power source or load. Energy is imparted to the system to start the oscillations, but thereafter energy neither enters nor

leaves the resonator. However, the energy does shift back and forth continuously between the inductor and capacitor, all of it being in L when e is zero and all in C when i is zero. The continuous oscillation of a constant amount of energy between electric-field energy (energy stored in a capacitor) and magnetic-field energy (energy stored in an inductor) is characteristic of all types of lossless resonators. The *amount* of energy involved depends upon the *amplitude* of the oscillations. Thus stored energy is a useful measure of the amplitude. From the energy stored in an oscillation, the peak value of each voltage and current in the oscillation may be found.

A resonator having two natural modes is illustrated in Fig. 2*A*. The circuit is drawn so that it resembles a two-section artificial transmission

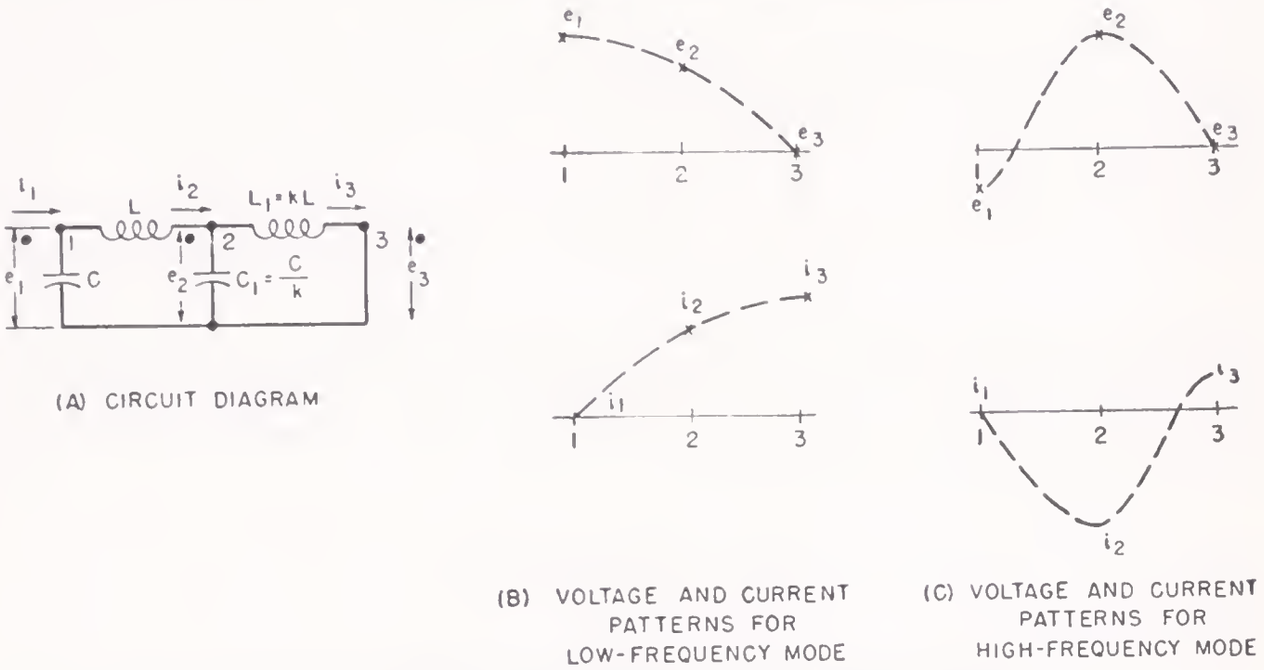


FIG. 2. Resonator having two natural frequencies.

line short-circuited at one end ($e_3 = 0$) and open-circuited at the other ($i_1 = 0$). Below f_0, the frequency at which L and C and also L_1 and C_1 resonate, the reactance of the parallel combination of L_1 and C_1 is inductive. The L_1, C_1 combination in series with L may therefore be thought of as an inductance that resonates with C at some frequency less than f_0. Similarly, at frequencies greater than f_0, the L_1, C_1 combination has a capacitive reactance, and the equivalent capacitance of the L_1, C_1 combination and C in series is less than C and resonates with L at a frequency greater than f_0.

Actual values of the two natural frequencies depend, of course, upon the value of k (see Fig. 2*A*). A convenient value for illustration is $k = 9/4$; the natural frequencies are then $f_1 = \frac{1}{2}f_0$ and $f_2 = 2f_0$. If the magnitude of the reactance of either L or C at f_0 is denoted by X, then at $f_1 = \frac{1}{2}f_0$ the reactance of C is $2X$ and that of L is $\frac{1}{2}X$. The sum of the inductive reactances of L and of the parallel combination of L_1 and C_1 is

$$\frac{X}{2} + \frac{1}{\frac{8}{9X} - \frac{2}{9X}} = \frac{X}{2} + \frac{3X}{2} = 2X$$

Thus at frequency f_1 the total inductive reactance of the circuit is equal to the capacitive reactance, and a natural mode occurs. Similarly, at $f_2 = 2f_0$ the reactance of L is $2X$, the reactance of C is $\frac{1}{2}X$, and the capacitive reactance of L_1 and C_1 in parallel is $\frac{3}{2}X$. The total capacitive reactance is thus equal to the inductive reactance, and f_2 is also a natural frequency.

In order to specify the mode patterns, it is necessary to determine the relative magnitudes and phases of voltages e_1, e_2, and e_3 and of currents i_1, i_2, and i_3. For oscillation in the low-frequency mode, i_2 flows through L and the equivalent inductance of the parallel combination of L_1 and C_1. Thus e_1 and e_2 are in phase, and since the reactance of L is $\frac{1}{2}X$ and the reactance of L_1 and C_1 in parallel is $\frac{3}{2}X$, at any instant e_2 and e_1 are in the ratio

$$\frac{e_2}{e_1} = \frac{3X/2}{2X} = \frac{3}{4}$$

Similarly, currents i_2 and i_3 result from the application of voltage e_2 across the parallel combination of L_1 and C_1 and across L_1, respectively. Thus at any instant i_2 and i_3 are in phase and have magnitudes in the ratio of the inductive reactance

$$\frac{i_2}{i_3} = \frac{9X/8}{3X/2} = \frac{3}{4}$$

The relative magnitudes of voltages and currents at points 1, 2, and 3 of the network are plotted in Fig. 2*B*. The plotted points are connected by dashed lines to suggest a resemblance between the network and a quarter-wave section of shorted-end transmission line. All the voltages are in phase, as are all the currents, but each current lags 90 deg behind the corresponding voltage.

For the high-frequency mode, voltage e_2 appears across a capacitive reactance of magnitude $\frac{3}{2}X$, and e_1 is across an inductive reactance of $2X - \frac{3}{2}X = X/2$. Thus e_1 and e_2 are in opposite phase, and their magnitudes are in the ratio of $\frac{1}{2}:\frac{3}{2}$, or 1:3. The voltage pattern of the mode is shown in Fig. 2*C*, the out-of-phase relation of e_1 and e_2 being represented by plotting magnitudes on opposite sides of the zero-voltage axis. The corresponding current pattern is obtained by noting that voltage e_2 across an inductive reactance causes current i_3 and the same voltage across the capacitive reactance of L_1 and C_1 in parallel causes current i_2. Thus i_2 and i_3 have opposite phase, and their magnitudes are in the ratio 3:1, which is the inverse ratio of the reactances. The currents

are 90 deg out of phase with the voltages. A plot of the current pattern is shown in Fig. 2C. The voltage and current patterns of the high-frequency mode again suggest a resemblance of the network to a transmission line—this time to a line in three-quarter-wave resonance.

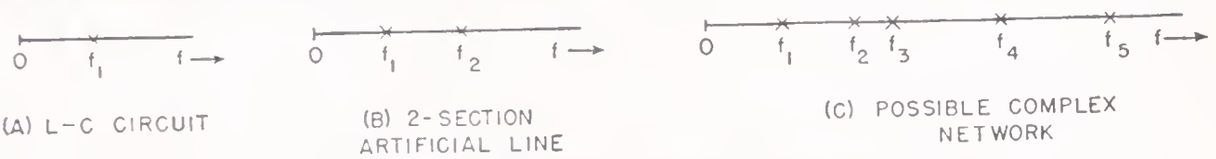
Knowledge of the voltage and current patterns of a mode and of the energy stored in an oscillation in that mode makes possible the calculation of the magnitude of any voltage or current produced by the oscillation. Suppose that the network of Fig. 2A is oscillating in the high-frequency mode and that the stored energy of the oscillation is W. From the mode pattern it is possible, for example, to calculate the relation of I_{m2}, the peak value of current i_2, to W. Because of the 90-deg phase relation between voltages and currents, all the stored energy is in the inductances at the instant of peak currents. Thus, if I_{m3} is the peak value of i_3,

$$W = \tfrac{1}{2}LI_{m2}^2 + \tfrac{1}{2}L_1I_{m3}^2 = \tfrac{1}{2}LI_{m2}^2 + (\tfrac{1}{2})(\tfrac{9}{4}L)(\tfrac{1}{3}I_{m2})^2 = \tfrac{5}{8}LI_{m2}^2$$

or $I_{m2} = \sqrt{8W/5L}$. If a peak voltage rather than a peak current is desired, the stored energy is expressed in terms of capacitances and voltages, and the desired peak voltage is obtained from the resultant relationship.

Because every resonant mode has definite voltage and current patterns, any voltage or current magnitude or the total stored energy may be used as a measure of the amplitude of oscillation in the mode. The stored energy does not refer to any special point in the network and is thus an especially useful measure of the oscillation amplitude.

A convenient way of specifying the resonant frequencies of a network is in terms of its *mode spectrum*, which is a plot of the natural frequencies of the network on a scale of frequency. The tuned circuit of Fig. 1 has the mode spectrum drawn in Fig. 3A, and the two-section artificial line,



(A) L-C CIRCUIT (B) 2-SECTION ARTIFICIAL LINE (C) POSSIBLE COMPLEX NETWORK

FIG. 3. Mode spectra of reactive networks.

the spectrum of Fig. 3B. More complicated networks in general have many natural frequencies, a possible spectrum being that of Fig. 3C.

2. The Quality Factor Q. Because networks always have some losses, the stored energy in any mode does not remain constant but gradually decreases as it is consumed by the dissipative elements. In practical resonators, the amount of energy dissipated per cycle of oscillation is extremely small compared with the energy stored in the system. The power loss and the resultant rate of energy decay may therefore be determined by assuming that the voltage and current patterns of the lossless resonator remain undisturbed in the presence of losses and by calculating

the losses produced by these voltages and currents in the dissipative elements. This method of accounting for small losses in a network is similar to the perturbation method used in the preceding chapters to determine the attenuation constants of transmission lines and waveguides (Art. 13, Chap. VII, and Art. 9, Chap. VIII).

A useful figure of merit of a network as a resonator is the ratio of the energy stored in the network to the energy dissipated in a specified time interval. If the time interval is $1/2\pi$ times the period of oscillation, this ratio is the *quality factor* Q of the resonator defined in Art. 18 of Chap. II as

$$Q = 2\pi \frac{\text{stored energy}}{\text{energy lost per cycle of oscillation}} \tag{5}$$

A relation equivalent to Eq. 5 is

$$Q = \omega_0 \frac{W}{P} \tag{6}$$

where ω_0 is the angular frequency of the oscillation, W is the stored energy, and P is the power loss.

As an example of the calculation of Q, consider the resonator of Fig. 1 and assume that both L and C have small losses which are accounted for by R_L and R_C in Fig. 4. In absence of losses $E_m = \sqrt{L/C}\, I_m$, $\omega_0 = 1/\sqrt{LC}$, and the stored energy is

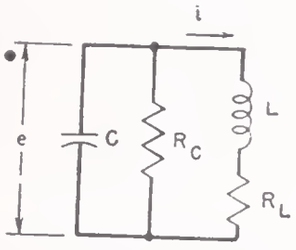


FIG. 4. The resonator of Fig. 1 with losses included.

$$W = \tfrac{1}{2}CE_m^2 = \tfrac{1}{2}LI_m^2$$

On the assumption that E_m and I_m are unaffected by the presence of losses, the power loss P in the network is

$$P = \frac{1}{2} I_m^2 R_L + \frac{1}{2} \frac{E_m^2}{R_C}$$

and

$$\frac{1}{Q} = \frac{P}{\omega_0 W} = \frac{\tfrac{1}{2}I_m^2 R_L + \tfrac{1}{2}(E_m^2/R_C)}{(1/\sqrt{LC}) \times \tfrac{1}{2}LI_m^2} = \frac{R_L}{\sqrt{L/C}} + \frac{\sqrt{L/C}}{R_C} \tag{7}$$

It is significant to note from the definition of Q that, because the stored energy is independent of the number of dissipative elements, the reciprocal of Q with several sources of loss present may be obtained by adding the reciprocals of the Q's which result when the sources of loss are inserted into the network one at a time. Thus $R_L/\sqrt{L/C}$ in Eq. 7 represents the reciprocal of the Q of Fig. 4 with only R_L present, and $\sqrt{L/C}/R_C$ is the reciprocal of the Q with only R_C present.

From Eq. 6 the power dissipated in a low-loss resonant network, which is the rate of decrease of stored energy, is

$$P = -\frac{dW}{dt} = \frac{\omega_0}{Q} W$$

Thus, because ω_0 and Q are constants, a curve of stored energy as a function of time has its slope proportional to its ordinate. The stored energy therefore decays exponentially (see Art. 4 of Chap. II) in accordance with the relation

$$W = W_o \epsilon^{-(\omega_0/Q)t} \tag{8}$$

where W_o is the amount of energy stored at $t = 0$ and ω_0/Q is the time constant of the exponential decay.

Because the peak value E_m of any voltage in the network (or the peak value I_m of any current) is proportional to the square root of the stored energy, the amplitudes of the voltage and current oscillations decrease exponentially. For a voltage,

$$E_m = E_{m0} \epsilon^{-(\omega_0/2Q)t}$$

where E_{m0} is the value of E_m at $t = 0$. In terms of instantaneous values,

$$e = E_{m0} \epsilon^{-(\omega_0/2Q)t} \cos \omega_0 t \tag{9}$$

3. Transmission-line Resonators. Sections of open- and shorted-end transmission lines useful as resonators are shown in Fig. 5*A*. The sec-

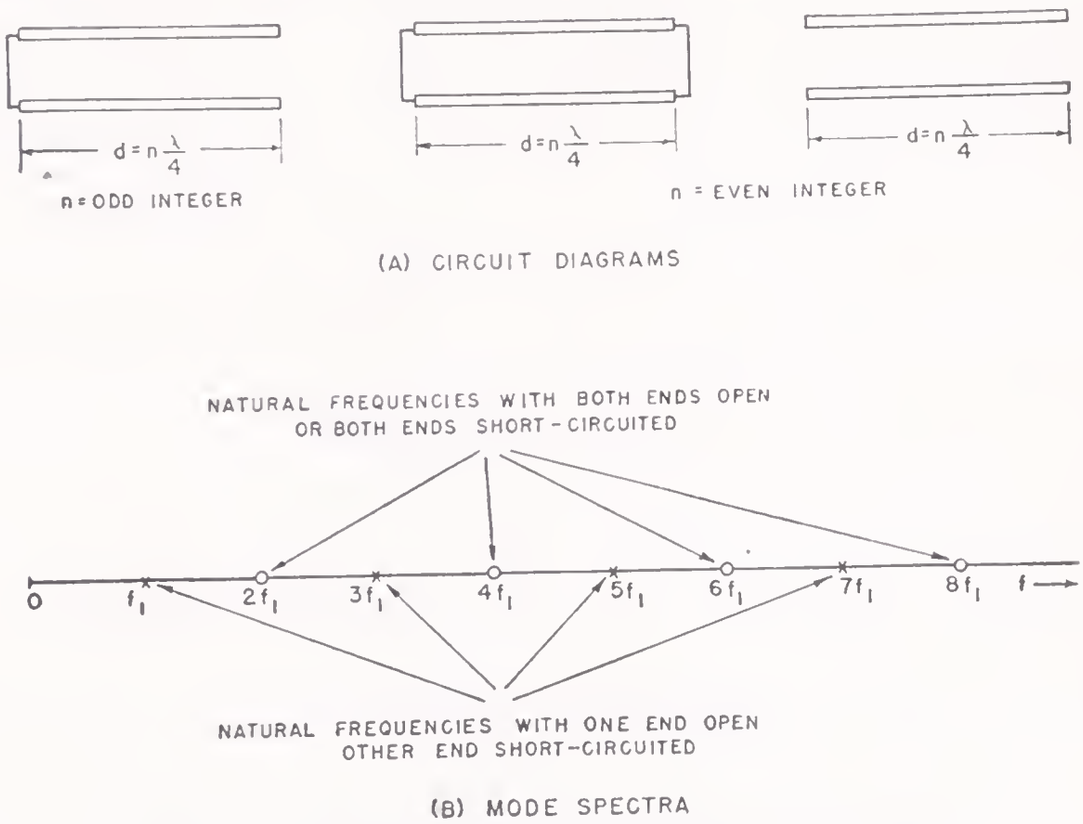


FIG. 5. Resonant line sections and their mode spectra.

tions are resonant at frequencies for which traveling waves of voltage and current, after making one round trip on the line and being reflected from the starting end, are exactly like the original waves delayed an integral

number of cycles. A source is necessary only during the first round trip of the waves, wave motion thereafter being sustained by successive reflections from opposite ends of the line.

On a line shorted at one end and open at the other, the total phase change in a single round trip of a wave is 360 deg (or a multiple thereof) at frequencies for which the line length is n quarter wavelengths, n being an odd integer. Thus if $n = 1$, the phase of a voltage wave changes 90 deg in traveling from the open to the shorted end, 180 deg at the short circuit, and another 90 deg in returning to the open end. Since reflection at the open end is without phase reversal, the total phase change is 360 deg. Similarly, the accompanying current wave is delayed 90 deg in traveling to the shorted end, another 90 deg in returning to the open end, and 180 deg as the result of reflection at the open end—again a total of 360 deg. If f_1 is the lowest natural frequency, corresponding to the quarter-wave mode, resonances occur at f_1, $3f_1$, $5f_1$, and so on, as indicated by the cross marks in the mode spectrum of Fig. 5*B*. In terms of T, the one-way delay time of the line, $f_1 = 1/4T$. If l and c are the inductance and capacitance per unit length of line and d is the line length, $T = d\sqrt{lc}$ and $f_1 = 1/4d\sqrt{lc}$.

When both ends of a line section are either open or shorted, resonances occur at frequencies for which the line length is n quarter wavelengths, where n is an even integer. The natural frequencies are thus $2f_1$, $4f_1$, $6f_1$, and so on, as marked by circles in Fig. 5*B*.

Frequently a tuning capacitor is connected at the open end of a resonant-line section, as in Fig. 6*A*, so that the natural frequencies may

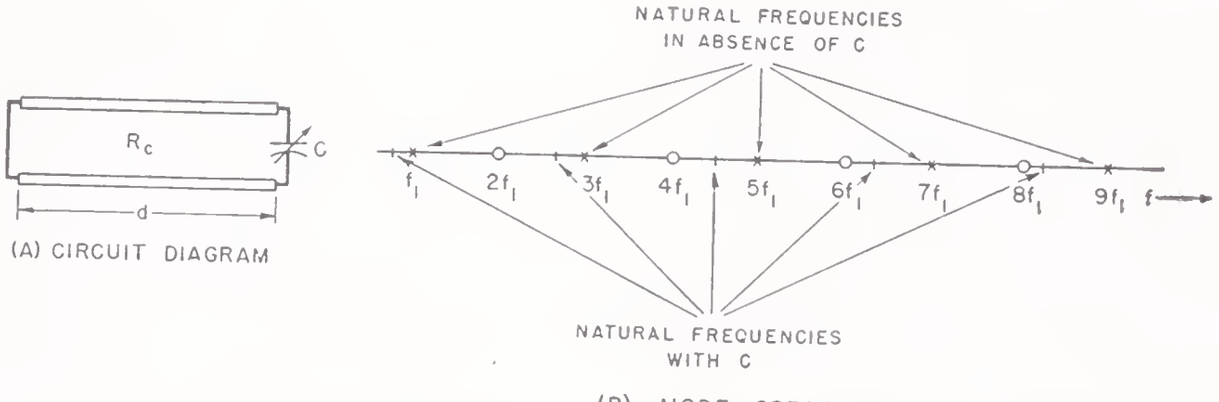


FIG. 6. Shorted-end line tuned to resonance by means of a capacitor.

be changed without changing the physical length of the line. Resonances then occur at frequencies for which the line has an inductive reactance $R_c \tan \beta d$ equal to the capacitor reactance $1/\omega C$. Thus at resonance, $\tan \beta d = 1/\omega R_c C$. If C is zero, $\tan \beta d = \infty$, and the natural frequencies occur for βd equal to an odd multiple of $\pi/2$, or d an odd multiple of $\lambda/4$, that is, at f_1, $3f_1$, $5f_1$, and so on as in Fig. 5*A*. In the presence of C, however, resonances occur for smaller values of βd, and the mode spectrum is shifted in the manner shown in Fig. 6*B*. The shift in natural fre-

quencies is more pronounced at the high-order modes than at the low-order modes because $1/\omega R_c C$ is inversely proportional to frequency. At frequencies far above f_1, C is nearly a short circuit, and the line resonates almost as though both ends were shorted.

An alternate method of adjusting the natural frequencies of a parallel-wire line makes use of a movable shorting bar as in Fig. 7*A*. If the bar is truly a short circuit, adjustments merely control the length of a line open at one end and shorted at the other, and the natural frequencies are inversely proportional to the length d. Because shorting bars always have some inductive reactance (particularly if their length is appreciable), the resonant device is actually composed of two resonators coupled by an inductance L, as shown in Fig. 7*B*. The resultant mode spectrum is

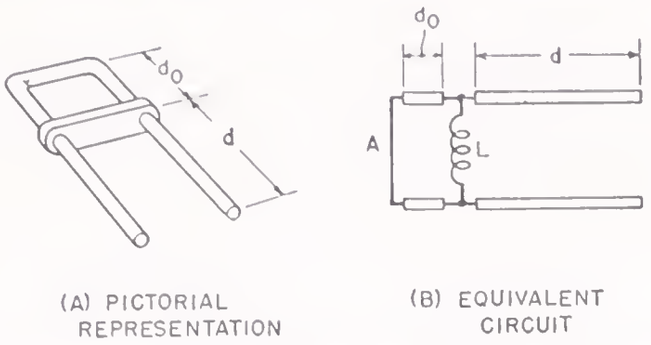


FIG. 7. Shorted-end line tuned to resonance by means of a shorting bar.

rather complicated, being dependent upon L and the relative lengths d and d_o. Usually, however, only the lowest natural frequency of the portion of line between shorting bar and open end is of interest. This section can be made to resonate in the desired manner by making distance d_o approximately $\lambda/4$ or less, so that the remaining line is terminated by an inductance only slightly different from L. Because the section of interest is terminated by an inductance, the lowest natural frequency occurs for a value of d slightly less than $\lambda/4$.

Observe that a short circuit, rather than an open circuit, is desirable at point A. Length d_o can then have any value less than about $\frac{3}{8}\lambda$. If an open circuit is employed at A, d_o must be much less than $\lambda/4$ in order to avoid placing sufficient capacitance in parallel with L to cause parallel resonance.

The mode pattern for a line in quarter-wave resonance is shown in Fig. 8*A*. The derivation of this pattern is given in the chapter on transmission lines (see Art. 3 of Chap. VII). As indicated in Fig. 8*A*, E_m, the peak value of the voltage along the line, varies in the manner of a sine wave, being zero at the shorted end and a maximum $E_{m_{\max}}$ at the open end. The peak value I_m of the current decreases as a cosine wave from $I_{m_{\max}}$ at the short circuit to zero at the open end. Thus $E_m = E_{m_{\max}} \sin \pi s/2d$ and $I_m = I_{m_{\max}} \cos \pi s/2d$ specify the peak value of the voltage and current along the line. (Compare these two relations with Eqs. 21 and 22 in

Chap. VII. Note that $\beta = \pi/2d$ for the quarter-wave resonance.) All the voltages are in phase, all the currents are in phase, and at each point of the line the phase angle between voltage and current is 90 deg. In accordance with Eqs. 21 and 22, pp. 483 and 485, the ratio $E_{m_{max}}/I_{m_{max}}$ is equal to R_c, the characteristic resistance of the line.

The patterns of higher order modes of the line of Fig. 8 may be obtained by cascading quarter-wave resonance patterns. Thus the mode pattern of a line resonating so that its electrical length is $\frac{3}{4}\lambda$ is the pattern drawn

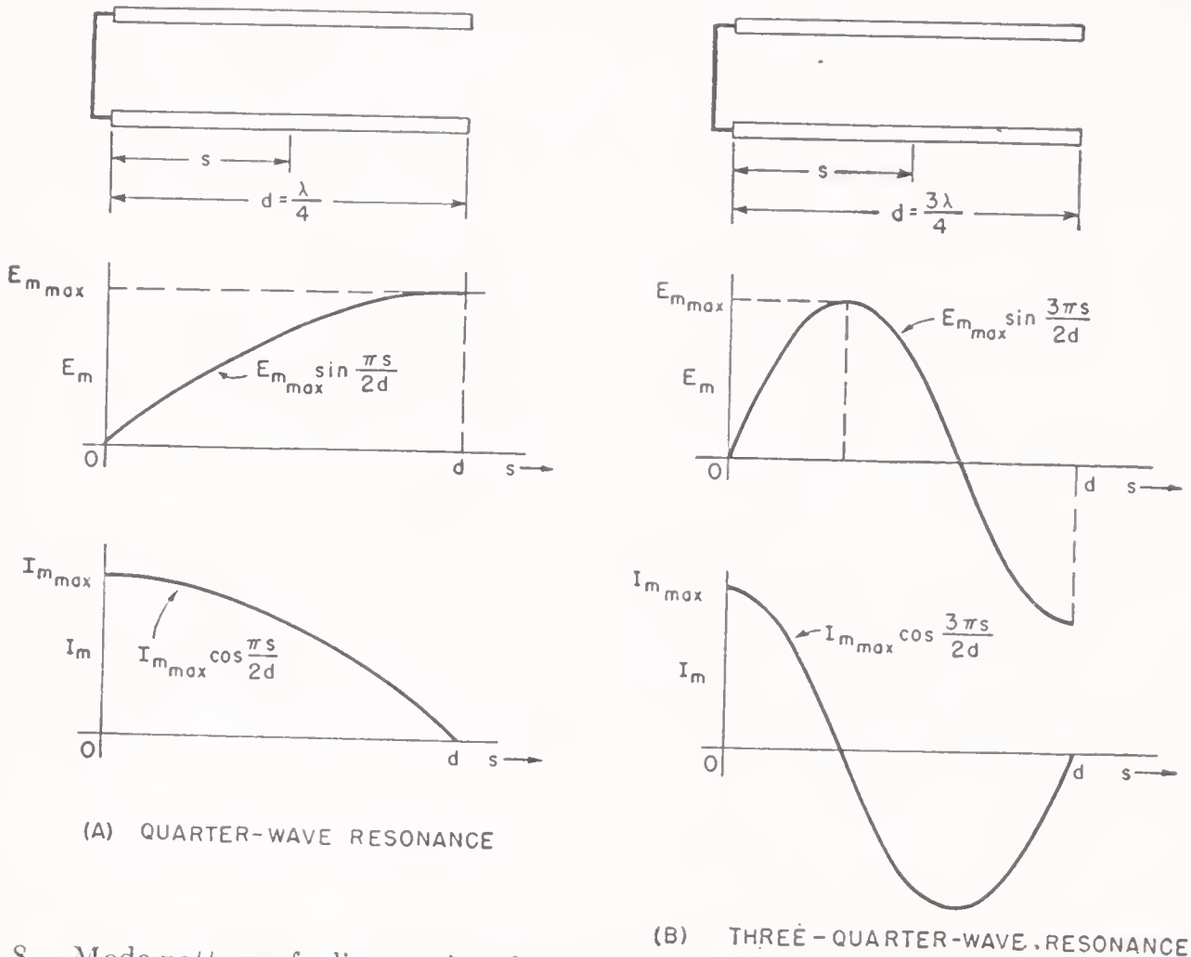


FIG. 8. Mode pattern of a line section short-circuited at one end and open at the other end.

in Fig. 8*B* and consists of three quarter-wave patterns joined to form continuous sine and cosine variations. Voltages and currents below the horizontal axes are 180 deg out of phase with the corresponding quantities plotted above the axes.

If a line short-circuited at one end is only a fraction of a quarter wave long and is resonating with a capacitor connected to the other end, the mode patterns are similar to those of a quarter-wave line except that they terminate before the full quarter-wave patterns are completed (see Fig. 9). Thus, if the electrical length of the line is βd, the maximum observable peak voltage is $E_{m_{max}} \sin \beta d$, but the maximum value of peak current is $I_{m_{max}}$, as on the quarter-wave line.

The calculation of stored energy in a resonant line is similar to that for a lumped-constant resonator, except that the sine-wave variations in peak current and peak voltage must be taken into account. Because of

the 90-deg phase relation between E_m and I_m, all the energy is in the magnetic field associated with the line inductance at instants of peak current and in the electric field associated with the line capacitance at instants of peak voltage. Thus stored energy may therefore be expressed either in terms of the peak voltage or peak current on the line. As an example of the calculation, consider the line of Fig. 9 to be resonating with C_0 in the lowest frequency mode. At an instant of peak current, the energy stored in a small line section of length Δs is

$$\Delta W = \frac{1}{2}(l\Delta s)(I_{m_{\max}} \cos \beta s)^2 = \frac{1}{2}\frac{l}{\beta} I_{m_{\max}}^2 \cos^2 \beta s \, \Delta(\beta s)$$

where l is the inductance per unit length of line and $\Delta(\beta s)$ is the increment of βs corresponding to Δs. Since the energy in C_0 is zero when the line current is maximum, the total stored energy W is the sum of the energies in all the $\Delta(\beta s)$ sections comprising the line. Thus

$$W = \frac{lI_{m_{\max}}^2}{2\beta} \int_0^{\beta d} \cos^2 \beta s \, d(\beta s)$$

$$= \frac{lI_{m_{\max}}^2}{2\beta} \int_0^{\beta d} \left(\frac{1 + \cos 2\beta s}{2}\right) d(\beta s)$$

Integration of this expression and substitution of limits yields

$$W = \frac{1}{2}(ld) I_{m_{\max}}^2 \left(\frac{1 + \dfrac{\sin 2\beta d}{2\beta d}}{2}\right) \quad (10)$$

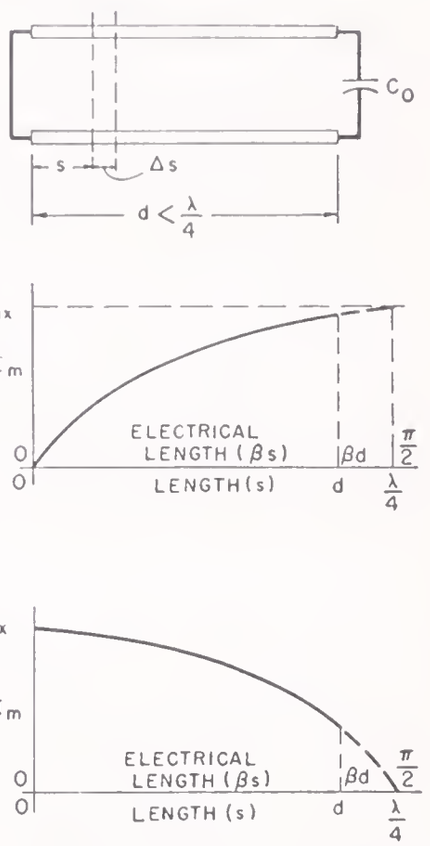


FIG. 9. Mode pattern of a line section with a short circuit at one end and a capacitor at the other end.

The factor $\frac{1}{2}(1 + \sin 2\beta d/2\beta d)$ in Eq. 10 accounts for the difference between the energy stored in a lumped-constant tuned circuit having an inductance $L = ld$ and a peak current $I_{m_{\max}}$ and the energy stored in a shorted-end line of inductance ld and peak current $I_{m_{\max}}$ in the shorting bar. If the oscillation frequency is low so that βd is very small, $\frac{1}{2}(1 + (\sin 2\beta d)/2\beta d)$ is very nearly unity, and the stored energy is essentially $\frac{1}{2}(ld)I_{m_{\max}}^2$, which is the energy stored in the equivalent tuned circuit. If the oscillation frequency is such as to permit quarter-wave resonance ($\beta d = \pi/2$), the stored energy becomes $\frac{1}{4}(ld)I_{m_{\max}}^2$. Thus, because of the sine distribution of peak current on the line, the stored energy is reduced by a factor of $\frac{1}{2}$ from the value it would have were the distribution uniform.

It is of interest to observe that a line of length d resonating alone ($C_0 = 0$) stores $\frac{1}{4}(ld)I^2_{m\max}$ units of energy in all modes of oscillation, provided the peak current $I_{m\max}$ is the same for all modes. For example, if the oscillation frequency makes $\beta d = \frac{3}{2}\pi$, the mode patterns are those of Fig. 8*B* and each third of the line stores $\frac{1}{4}(ld/3)I^2_{m\max}$ units of energy. The total energy stored is therefore $\frac{1}{4}(ld)I^2_{m\max}$, which is the same as for quarter-wave resonance. Similarly, if both ends of the line are short-circuited and the mode of oscillation corresponds to two quarter-wave patterns of voltage and current ($\beta d = \pi$), each half of the line stores $\frac{1}{4}(ld/2)I^2_{m\max}$ units of energy and the total energy is again the same as for the quarter-wave line.

The quality factor Q of a resonant line may be determined from the ratio of the energy stored in the line to the power dissipated. In a small section Δs for which the peak current I_m may be considered constant, the stored energy is $\Delta W = \frac{1}{2}(l\Delta s)I_m^2$, and the power loss, if caused solely by the resistance of the line conductors, is $\Delta P = \frac{1}{2}(r\Delta s)I_m^2$, where r is the resistance per unit length of line. Thus the ratio of stored energy to power dissipated for the Δs section is

$$\frac{\Delta W}{\Delta P} = \frac{\frac{1}{2}(l\Delta s)I_m^2}{\frac{1}{2}(r\Delta s)I_m^2} = \frac{l}{r}$$

This ratio applies for each small section of the line. Therefore, the ratio of total stored energy to total power dissipation in the line is also l/r, and by Eq. 6, p. 618,

$$Q = \omega_0 \frac{W}{P} = \frac{\omega_0 l}{r} = \frac{\omega_0 L}{R} \tag{11}$$

where $L = ld$ and $R = rd$ are the total inductance and resistance of the line. This equation applies to a resonant line comprising any integral number of quarter-wave sections. For lines terminated in inductors or capacitors, the Q is influenced by the dissipation in the terminating elements.

The Q can also be related to α, the attenuation constant of the line. Since $\omega_0 = 2\pi f_0 = 2\pi v/\lambda$, and v, the speed of propagation of waves on the line, equals $1/\sqrt{lc}$,

$$Q = \frac{2\pi}{\lambda\sqrt{lc}}\frac{l}{r} = \frac{2\pi}{\lambda}\frac{\sqrt{l/c}}{r} = \frac{2\pi R_c}{\lambda r}$$

where $\sqrt{l/c} = R_c$, the characteristic resistance of the line. According to Eq. 60 of Chap. VII, p. 517, for a line having negligible dielectric loss $\alpha = r/2R_c$. Hence,

$$Q = \frac{\pi}{\lambda\alpha} \tag{12}$$

or Q is equal to π divided by the attenuation in nepers of a one-wave length of line.

In the above relations for Q, losses in the dielectric of the line, losses in short-circuiting bars, and radiation losses are ignored. Because of these losses, the actual Q is always lower than the value calculated from Eq. 11. The relative importance of losses at the line terminations may be reduced and Q increased by an increase of the physical length of the line so as to increase the stored energy. This result is one example of a general principle discussed in Art. 4: The Q of a resonant device improves with increase in size of the device. This principle may be illustrated in another way. If the cross-sectional dimensions of the line are increased in a manner that maintains R_c constant but increases the conductor size, the attenuation $\alpha = r/2R_c$ is decreased because of the decrease of r and the Q is increased.

As an example of the order of magnitude of Q to be expected for a resonant line, consider the Q of the line in Fig. 10. If the effect of the shorting bar is ignored, the quarter-wave resonance of the 5.9-in. line is found to occur at 500 Mcps. According to Eq. 88 of Chap. VII, p. 539, the inductance per unit length of line is $l = 0.922$ μh per m. For copper conductors of resistivity equal to 1.72×10^{-8} ohm-meter, the resistance per unit length, obtained from Eq. 91, p. 541, is $r = 0.292$ ohm per m. The Q of the line by Eq. 11 is therefore 9,920. Because of radiation from the ends of the line and losses in the short circuit, the actual Q is less than the calculated value, but it is nevertheless very high in comparison with the Q of a lumped-parameter circuit of equal resonant frequency.

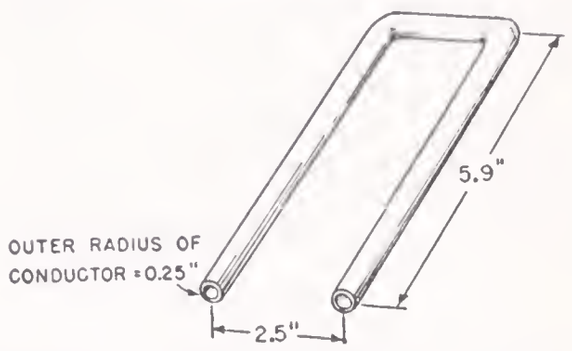


FIG. 10. Parallel-wire line used as example for calculation of Q.

Although the properties of resonant transmission lines have been discussed in this article with special reference to parallel-wire lines, the same principles apply also to coaxial lines. Various forms of coaxial-line resonators are described in Art. 5.

4. Rectangular-waveguide Resonators. Resonators for use in the microwave range of frequencies take the form of totally enclosed metallic chambers, or *cavities*. The cavity walls are made of a highly conducting material and enclose a good dielectric, usually air. One example of a cavity resonator is the rectangular box of Fig. 11*A*, which may be thought of as a section of rectangular waveguide closed at both ends by conducting plates. Because the end plates are short circuits for waves traveling in the z direction, the cavity is analogous to a transmission-line section with short circuits at both ends, and resonant modes occur at frequencies for

which the distance between end plates is a multiple of half the guide wavelength. Higher order waves, as well as the dominant wave, may give rise to resonant frequencies, and thus a great variety of resonant modes is possible. The mode spectrum is more complicated than that discussed for the shorted-end transmission line in Art. 3.

The characteristics of interest for each mode of a cavity resonator are the mode pattern, natural frequency, stored energy, and quality factor Q —the same characteristics considered in connection with network and transmission-line resonators. In the discussion of these characteristics for the modes of the resonator of Fig. 11*A* of dimensions a, b, and d as

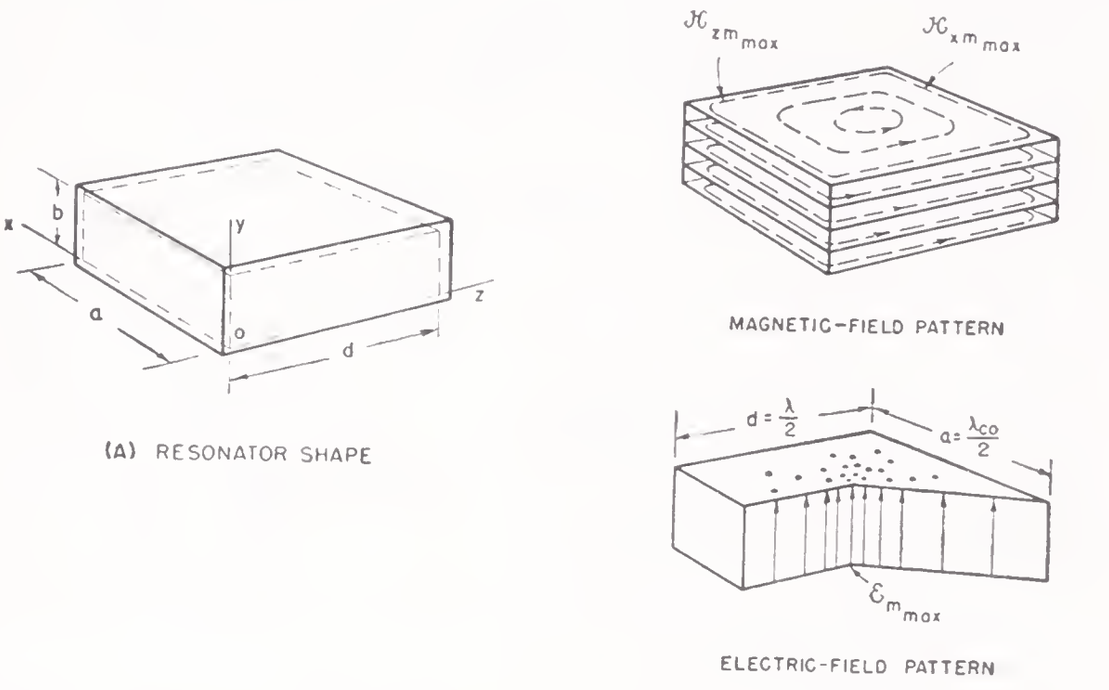


FIG. 11. Rectangular-waveguide resonator.

indicated, special reference is made to a particularly simple mode—the mode obtained when $TE_{1,0}$ waves travel in the z direction and have a half wavelength equal to d.

The pattern of this mode is described by the field configurations of Fig. 11*B*. This figure is obtained by addition of the fields of the component traveling waves. Observe that the pattern is like one segment of the pattern of the $TE_{1,0}$ wave (Fig. 13, of Chap. VIII, p. 573), except that the loops formed by the magnetic-field lines are shifted one-quarter wavelength with respect to the electric-field lines. This shift is the result of adding the two oppositely directed traveling waves. The electric-field intensities add to zero at the end plates and to a maximum of twice the field intensity of the component wave at the center, whereas the transverse-magnetic-field intensity doubles at the end plates and becomes zero at the center. Observe also that throughout the cavity the electric field has the same phase. Similarly the magnetic field increases and

decreases as a unit, but the variations of the magnetic field are 90 deg out of phase with those of the electric field. As in other resonant systems, the energy of the oscillations remains constant in total value as it shifts back and forth between electric- and magnetic-field forms.

Let $\mathcal{E}_{m_{\max}}$ be the maximum value (attained in the center of the cavity) of the peak electric-field intensity, and let $\mathcal{H}_{xm_{\max}}$ and $\mathcal{H}_{zm_{\max}}$ be the maximum values (attained near the vertical walls) of the x-directed and z-directed components of the peak magnetic-field intensity. Because these maximum quantities are each twice the corresponding quantities of the component traveling waves, the ratio of electric-field intensity to magnetic-field intensity is the same as in the traveling waves, and from Eqs. 44 and 45 of Chap. VIII, p. 575,

$$\mathcal{H}_{xm_{\max}} = \frac{\mathcal{E}_{m_{\max}}}{Z_i} \frac{\lambda_0}{\lambda} \tag{13}$$

and

$$\mathcal{H}_{zm_{\max}} = \frac{\mathcal{E}_{m_{\max}}}{Z_i} \frac{\lambda_0}{\lambda_{co}} \tag{14}$$

These equations together with Fig. 11*B* and the sinusoidal character of the field-intensity variations specify the mode pattern. They may be put in a more convenient form by use of the relations $\lambda = 2d$ and $\lambda_{co} = 2a$. Dividing Eq. 13 by Eq. 14 yields the ratio

$$\frac{\mathcal{H}_{xm_{\max}}}{\mathcal{H}_{zm_{\max}}} = \frac{a}{d} \tag{15}$$

Squaring Eqs. 13 and 14, adding, and then applying Eq. 42 of Chap. VIII, p. 575, yields

$$\mathcal{E}^2_{m_{\max}} = Z_i^2(\mathcal{H}^2_{xm_{\max}} + \mathcal{H}^2_{zm_{\max}}) \tag{16}$$

Modes in the rectangular cavity are designated in terms of the traveling waves that give rise to the modes. To the symbol for the traveling wave is added a subscript p to denote the number of half-cycle variations in the standing-wave pattern along the direction of wave travel in the guide. Thus the mode of Fig. 11*B*, considered to be formed by $TE_{1,0}$ waves traveling in the z direction with d equal to a half wavelength, is designated $TE_{1,0,1}$. Other $TE_{m,0,p}$ modes may be visualized in terms of Fig. 11*B* if the cavity is divided into rectangular cells, dimension a being split into m parts and d into p parts. Each cell contains a miniature reproduction of the pattern of Fig. 11*B*. For the more general $TE_{m,n,p}$ modes, n half-cycle variations along the b dimension must also be considered; for the $TM_{m,n,p}$ modes similar patterns apply, but the magnetic field forms closed loops in planes perpendicular to the z axis.

Complete specification of a mode requires that the assumed direction of

wave travel as well as the designation symbol be stated. For example, the $TE_{1,0,1}$ mode has the configuration of Fig. 11*B* if waves traveling in the z direction of Fig. 11*A* are considered. If, however, waves traveling parallel to the y axis are assumed, the $TE_{1,0,1}$ symbol designates a different mode having fields perpendicular to those of Fig. 11*B* and having a different natural frequency.

The natural frequency of any $TE_{m,n,p}$ or $TM_{m,n,p}$ mode may be obtained from the wavelength relations of Eq. 42, p. 575, and Eq. 62 of Chap. VIII, p. 590,

$$\frac{1}{\lambda_0^2} = \frac{1}{\lambda_{co}^2} + \frac{1}{\lambda^2}$$

$$\frac{1}{\lambda_{co}^2} = \left(\frac{m}{2a}\right)^2 + \left(\frac{n}{2b}\right)^2$$

and the relation $d = p\lambda/2$. From these relations

$$\left(\frac{f}{v}\right)^2 = \frac{1}{\lambda_0^2} = \left(\frac{m}{2a}\right)^2 + \left(\frac{n}{2b}\right)^2 + \left(\frac{p}{2d}\right)^2$$

and

$$f = v\sqrt{\left(\frac{m}{2a}\right)^2 + \left(\frac{n}{2b}\right)^2 + \left(\frac{p}{2d}\right)^2} \tag{17}$$

where f is the natural frequency of the mode and v is the speed of the traveling waves and equals the speed of light in space if the dielectric is air. For the $TE_{1,0,1}$ mode of Fig. 11, the frequency is

$$f = v\sqrt{\left(\frac{1}{2a}\right)^2 + \left(\frac{1}{2d}\right)^2} = f_{co}\sqrt{1 + \left(\frac{a}{d}\right)^2}$$

where f_{co} is the cutoff frequency for $TE_{1,0}$ waves traveling parallel to the z axis.

The method of designating modes and a portion of the mode spectrum of a rectangular-waveguide resonator are illustrated in Fig. 12, which applies to a resonator for which $b = 1/\sqrt{5}$, $a = 1/\sqrt{4}$, and $d = 1/\sqrt{3}$. The mode designations in Fig. 12*B* refer to waves propagating parallel to the z axis in Fig. 12*A*, and the natural frequencies indicated are computed from Eq. 17, which for the dimensions of Fig. 12*A* may be written as

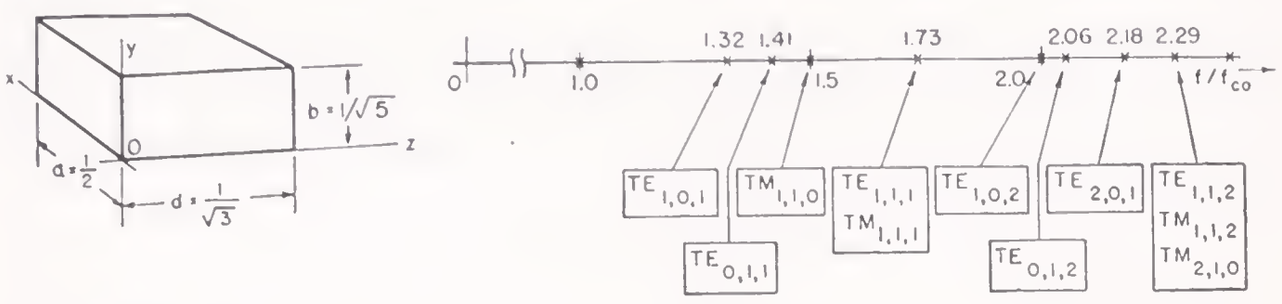
$$f = \tfrac{1}{2}f_{co}\sqrt{4m^2 + 5n^2 + 3p^2}$$

where f_{co} is the cutoff frequency for $TE_{1,0}$ waves traveling in the z direction.

Among the modes of Fig. 12*B* are two ($TM_{1,1,0}$ and $TM_{2,1,0}$) that arise from transverse-magnetic waves and for which the fields are independent of the z coordinate. These modes occur at the cutoff frequencies of the

associated transverse-magnetic waves and may be looked upon as the resultant of transverse-electromagnetic waves traveling in directions perpendicular to the z axis. (See the discussion in Arts. 7 and 12 of Chap. VIII of waveguide operation at the cutoff frequency.) Similar resonances at the cutoff frequencies of transverse-electric waves cannot occur because components of electric-field intensity along the conducting end plates would be required.

Also included in Fig. 12*B* are examples of *degenerate modes*, that is, modes made up of two or more field patterns having the same natural frequency. For the $TE_{1,1,1}$ and $TM_{1,1,1}$ resonances in Fig. 12*B*, the natural frequency is $1.73f_{co}$; and for the $TE_{1,1,2}$, $TM_{1,1,2}$, and $TM_{2,1,0}$ modes, the natural frequency is $2.29f_{co}$. The field pattern of a degenerate mode is



(A) DIMENSIONS AND COORDINATE AXES

(B) MODE SPECTRUM

FIG. 12. Portion of mode spectrum for a rectangular-waveguide resonator.

not definite, but it may take any of a variety of forms obtained by adding the fields of the component patterns in various amplitude and phase relationships. If the mode spectrum is determined with reference to waves traveling in the x or y direction in Fig. 12*A*, not only are the designation symbols for the modes changed, but also the component patterns in terms of which the degenerate modes are represented are altered.

The stored energy of network and transmission-line resonators has been calculated in Arts. 1 and 3 from the relations $\frac{1}{2}Ce^2$ and $\frac{1}{2}Li^2$, which express the energy in the field of a capacitor or an inductor. For use with cavity resonators, the analogous field-quantity relations $\frac{1}{2}\varepsilon\mathcal{E}^2$ and $\frac{1}{2}\mu\mathcal{H}^2$ are required. These express the *energy per unit volume* of an electric field of intensity $\mathcal{E}$ and a magnetic field of intensity $\mathcal{H}$, respectively. The field-quantity relations yield the same results as the relations $\frac{1}{2}Ce^2$ and $\frac{1}{2}Li^2$ when applied to compute capacitor or inductor energy. They have the advantage of being useful in other circumstances (for example, in waveguide and cavity-resonator calculations) where lumped inductance and capacitance elements associated with separate regions of space cannot be specified.

As an example of the equivalence of the circuit and field relations for energy, consider a capacitor formed of two parallel metal plates, each of area A separated by a dielectric of thickness d and permittivity ε, the

capacitance C being $\varepsilon A/d$. If the capacitor is charged, an electric field of intensity $\mathcal{E}$ is present in the dielectric and a voltage $\mathcal{E}d$ appears across the capacitor. The energy in the electric field, by the circuit relation, is

$$\frac{1}{2}Ce^2 = \frac{1}{2}\left(\varepsilon\frac{A}{d}\right)(\mathcal{E}d)^2 = \frac{1}{2}\varepsilon\mathcal{E}^2Ad$$

This result is in agreement with the field-quantity relation because $\frac{1}{2}\varepsilon\mathcal{E}^2$ is the energy per unit volume and Ad is the volume occupied by the field.

For any cavity resonator, the stored energy of an oscillation shifts back and forth between electric- and magnetic-field forms and is equal to the *peak* energy in either the electric or the magnetic field. Consider first the electric-field energy which has its peak value at an instant when the electric-field intensity at each point of the cavity has its peak value $\mathcal{E}_m$. The energy ΔW stored in a small volume Δv throughout which the field intensity is essentially uniform is $\frac{1}{2}\varepsilon\mathcal{E}_m^2\Delta v$, and the total energy W is the sum of the components ΔW taken over all the small volumes Δv in the cavity, or

$$W = \tfrac{1}{2}\varepsilon\int_V \mathcal{E}_m^2 dv \tag{18}$$

where the letter V indicates that the integration is to be performed over the entire volume of the cavity. In a similar way, if $\mathcal{H}_m$ is the peak magnetic-field intensity at each point of the cavity, the stored energy of the oscillation is

$$W = \tfrac{1}{2}\mu\int_V \mathcal{H}_m^2 dv \tag{19}$$

Equations 18 and 19 must necessarily yield the same result for any cavity.

The meaning of these energy relations may be illustrated by applying them to the rectangular-cavity mode of Fig. 11*B*. The electric-field intensity has a peak value $\mathcal{E}_{m_{\max}}$ at the center of the resonator and varies sinusoidally in the x and z directions; that is,

$$\mathcal{E}_m = \mathcal{E}_{m_{\max}}\sin\frac{\pi x}{a}\sin\frac{\pi z}{d}$$

and by Eq. 18 the stored energy is

$$W = \frac{1}{2}\varepsilon\int_V \mathcal{E}_{m_{\max}}^2\sin^2\frac{\pi x}{a}\sin^2\frac{\pi z}{d}\,dv$$

where dv can be taken as the volume of a vertical column in the resonator of height b and cross-sectional dimensions dx and dz. The integration introduces a factor of $\frac{1}{2}$ for each of the sine-squared factors (see Art. 9 of Chap. VIII), and thus the stored energy is

$$W = \tfrac{1}{8}\varepsilon\mathcal{E}_{m_{\max}}^2 V \tag{20}$$

If Eq. 19 rather than Eq. 18 is used to evaluate the stored energy, $\mathcal{H}_m^2$ may be replaced by $\mathcal{H}_{xm}^2 + \mathcal{H}_{zm}^2$. Because each of these terms has a sine-squared variation in two directions, the result of the integration is $\frac{1}{8}\mu(\mathcal{H}_{xm_{max}}^2 + \mathcal{H}_{zm_{max}}^2)V$. Equation 16 may be used to show that this value of stored energy is equal to the value given by Eq. 20.

The power loss in a cavity resonator may be calculated by a perturbation method very similar to the method of calculation of attenuation of a transmission line or waveguide (see Art. 13 of Chap. VII and Art. 9 of Chap. VIII). The power loss may ordinarily be considered to be entirely the result of currents in the resonator walls, and the average-power dissipation per unit area of wall surface is, by Eq. 55 of Chap. VIII, p. 584, $\frac{1}{2}\frac{\rho}{\delta}\mathcal{H}_m^2$, where $\mathcal{H}_m$ is the peak value of the magnetic-field intensity at the wall surface. The total power loss in the entire wall enclosing the resonator volume is the sum of the losses in many small areas comprising the wall surface and is given by the integral

$$P = \frac{1}{2}\frac{\rho}{\delta}\int_A \mathcal{H}_m^2 da \tag{21}$$

where da is an element of area and A signifies the entire wall-surface area over which the integration is to be performed.

By Eq. 6, p. 618, the quality factor Q is equal to $\omega_0 W/P$, and thus Eqs. 19 and 21 may be combined to yield

$$Q = \frac{\omega_0 W}{P} = \frac{\omega_0 \frac{1}{2}\mu \int_V \mathcal{H}_m^2 dv}{\frac{1}{2}\frac{\rho}{\delta}\int_A \mathcal{H}_m^2 da} = \frac{2\pi f_0 \mu}{\rho}\,\delta\,\frac{\int_V \mathcal{H}_m^2 dv}{\int_A \mathcal{H}_m^2 da}$$

By Eq. 89 of Chap. VII, p. 540, $\pi f_0 \mu_c/\rho = 1/\delta^2$; and therefore if μ and μ_c, the permeabilities of the dielectric and conductor materials, are equal, as they ordinarily are,

$$Q = \frac{2}{\delta}\frac{\int_V \mathcal{H}_m^2 dv}{\int_A \mathcal{H}_m^2 da} \tag{22}$$

The integrals in Eq. 22 may be evaluated easily if the equation is applied to a resonant mode of a rectangular-waveguide resonator. Consider, for example, the mode of Fig. 11*B*, and assume the resonator to be cubical, the length of each side being d. The surface integral in the denominator may be expressed as the sum of six integrals, one for each side of the resonator. For the sides perpendicular to the x axis

$$\mathcal{H}_m = \mathcal{H}_{z m_{max}}^2 \sin^2 \frac{\pi x}{a}$$

and thus (because the average value of the sine squared factor is ½) the integral over each side is $\frac{1}{2}\mathcal{H}^2_{zm\max}d^2$. Similarly the integral over the sides perpendicular to the z axis is $\frac{1}{2}\mathcal{H}^2_{xm\max}d^2$. For the sides perpendicular to the y axis, both components of the magnetic field are present, and each varies sinusoidally in two directions. Thus the integral over the top or the bottom surface is $\frac{1}{4}(\mathcal{H}^2_{xm\max} + \mathcal{H}^2_{zm\max})d^2$, and the total surface integral is

$$\begin{aligned}\int_A \mathcal{H}_m{}^2 da &= 2[\tfrac{1}{2}\mathcal{H}^2_{zm\max} + \tfrac{1}{2}\mathcal{H}^2_{xm\max} + \tfrac{1}{4}(\mathcal{H}^2_{xm\max} + \mathcal{H}^2_{zm\max})]d^2 \\ &= \tfrac{3}{2}(\mathcal{H}^2_{xm\max} + \mathcal{H}^2_{zm\max})d^2\end{aligned}$$

The volume integral of Eq. 22 was evaluated on p. 630 in the calculation of stored energy and is

$$\int_V \mathcal{H}_m{}^2 dv = \tfrac{1}{4}(\mathcal{H}^2_{xm\max} + \mathcal{H}^2_{zm\max})d^3$$

Therefore Q has the value

$$Q = \frac{2}{\delta}\frac{\frac{1}{4}(\mathcal{H}^2_{xm\max} + \mathcal{H}^2_{zm\max})d^3}{\frac{3}{2}(\mathcal{H}^2_{xm\max} + \mathcal{H}^2_{zm\max})d^2} = \frac{d}{3\delta} \tag{23}$$

The order of magnitude of Q to be expected from cavity resonators may be obtained by applying Eq. 23 to a cubical resonator adjusted to resonate in the $TE_{1,0,1}$ mode at 3,000 Mcps. From Eq. 17, p. 628, d is equal to $10/\sqrt{2} = 7.07$ cm or 2.78 in. If the resonator walls are copper, the skin depth is 1.21×10^{-6} m at 3,000 Mcps (see p. 540 of Chap. VII). The value of Q is therefore $7.07 \times 10^{-2}/3 \times 1.21 \times 10^{-6}$, or 19,500. This value of Q is much higher than could be obtained with a transmission-line resonator at the same frequency, but even higher Q values can be attained, for example, by silver-plating the resonator walls to reduce ρ or, as discussed in the next article, by employing the low-loss mode of a cylindrical resonator.

Equation 22 can be employed in the calculation of Q for cavity resonators of any shape, but its usefulness is limited because of the difficulty of evaluating the volume and surface integrals when the cavity shape or the field configuration is complicated. Several important general conclusions about the quality factor can be reached, however, without actually performing the integrations.

The right-hand member of Eq. 22 is the product of two factors. The first of these ($2/\delta$) indicates the effect of wall material and resonant frequency upon Q, and the second (the ratio of the integrals) shows the influence of the mode pattern, cavity size, and shape. Because $1/\delta = \sqrt{\pi f \mu_c/\rho}$, the first factor indicates that Q varies as the square root of the conductivity of the wall material and as the square root of fre-

quency. (Frequency changes require corresponding changes of the cavity size or shape, or of the mode pattern, and may thus have an additional effect on Q through changes in the second factor.) The second factor indicates a general tendency for Q to increase in proportion to the linear dimensions of the cavity—if the cavity shape and the mode pattern are fixed, the volume integral is proportional to the cube of the linear dimensions and the surface integral to the square of the same dimensions. The second factor shows also that for high Q a mode pattern is desirable in which the magnetic field in the interior of the cavity is as large as possible relative to that at the wall surfaces—the field in the interior of the cavity contributes to the volume integral, but not to the surface integral.

5. Cavity Resonators. The cavity formed by closing the ends of a rectangular waveguide is only one of many cavity devices useful as high-frequency resonators. By appropriate choice of cavity shape, advantages such as compactness, ease of tuning, simple mode spectrum, and high Q may be secured as required for special applications.

A section of coaxial line with a short-circuiting disk at each end, as in Fig. 13*A*, may be considered a form of resonant cavity. The coaxial cavity has a convenient size when designed for frequencies in the upper uhf and lower microwave ranges. Resonance occurs when the distance between short circuits is a multiple of a half wavelength for any wave type that may traverse the line. The field pattern for the lowest frequency mode—corresponding to half-wave resonance of the principal (transverse-electromagnetic) wave—is shown in Fig. 13*B*. Note that the directions of the electric-field lines have no significance relative to the directions of the magnetic-field lines; the two fields are 90 deg apart in phase.

The spectrum of the resonant modes produced by the principal wave in the coaxial cavity is the same as that described for parallel-wire lines in Art. 3 and comprises natural frequencies at integral multiples of the lowest natural frequency. Higher order transverse-electric and transverse-magnetic waves can propagate in coaxial lines and may produce additional modes of resonance at high frequencies. Often such modes are avoided by designing the line so that the cutoff frequency for all higher order waves is above the range of operating frequencies. This condition is obtained if the average circumference of the conductors is less than the shortest wavelength in the operating range.

As a means of tuning, a movable short-circuiting plunger is provided for the cavity as shown in Fig. 13*A*. The shaft that controls the plunger position may be threaded to permit fine control. A disadvantage of this tuning method is that the sliding contacts between the plunger and the inner and outer conductors may cause erratic operation and high losses, especially after the contact surfaces have suffered considerable wear through use.

In Fig. 13*C* is shown a noncontacting form of movable plunger designed to avoid the wear and losses associated with the contacting type. The plunger is a quarter-wave thick and forms the inner conductor of a very low-impedance coaxial line for which the resonator wall is the outer conductor. This line is terminated at *A* in an impedance much higher than its characteristic impedance and therefore has a very low input impedance at *B*. Thus, for high-frequency currents, the equivalent of a connection between plunger and outer wall is provided at *B*. Similarly, the low-

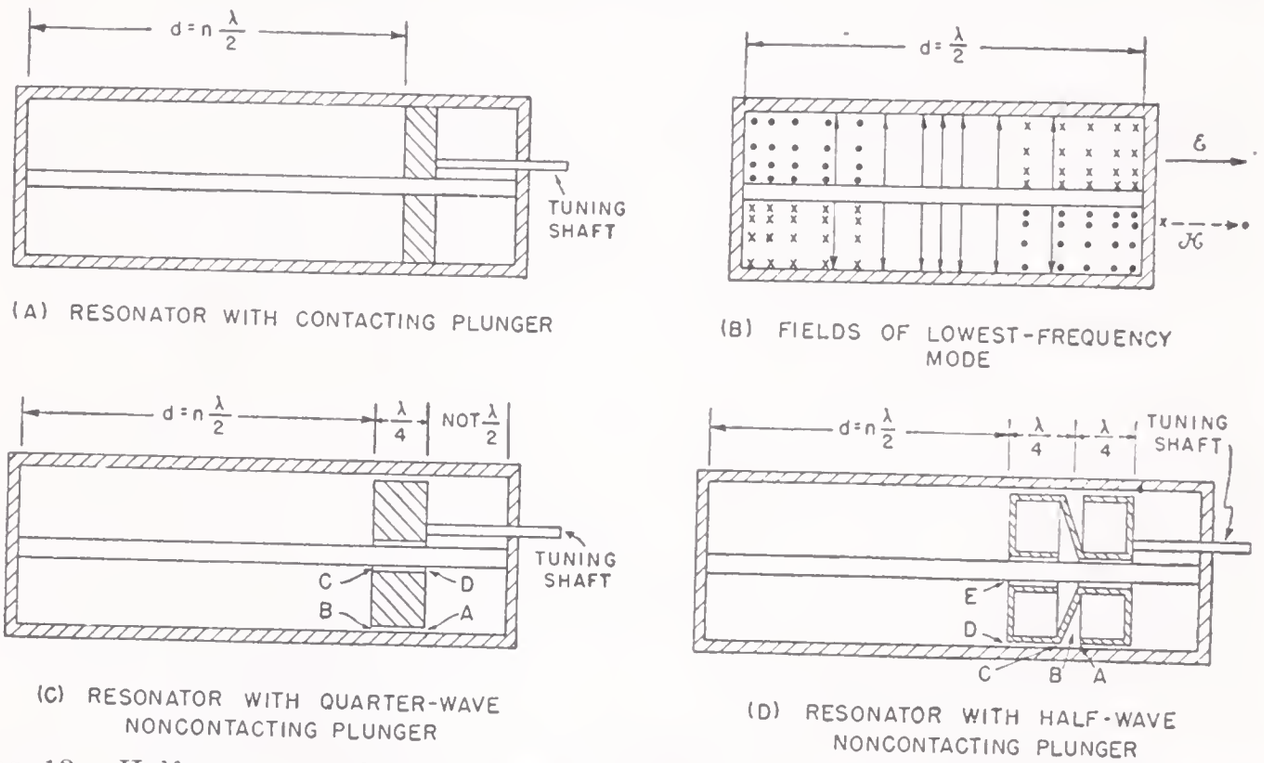


FIG. 13. Half-wave or multiple half-wave coaxial-line resonators shown in sectional view

impedance coaxial line between plunger and center rod provides an effective connection at point *C*.

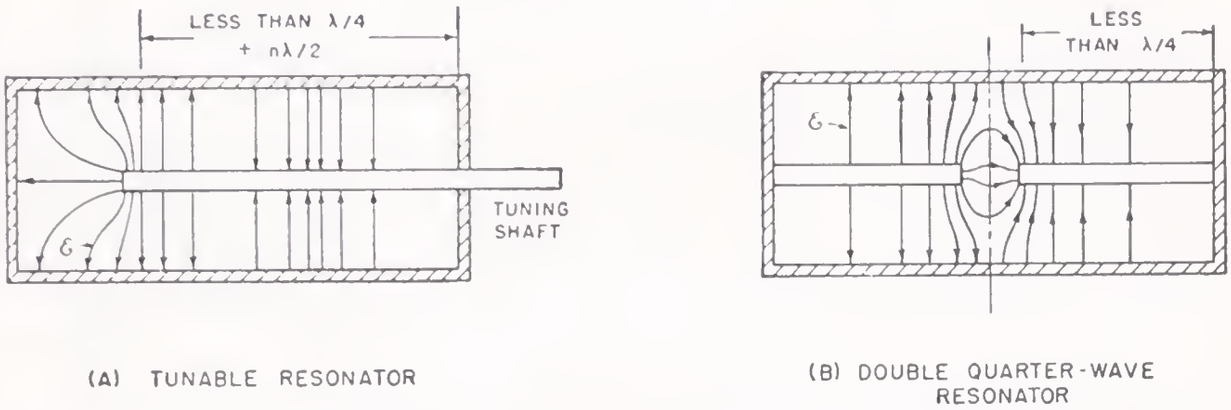
Two difficulties arise in the use of the plunger of Fig. 13*C*. One is that the tuning range is limited because the plunger thickness can approximate a quarter wave over only a relatively narrow frequency band. The other is that for certain critical plunger positions the cavity to the right of the plunger may be in multiple-half-wave resonance, thus presenting a low impedance at points *A* and *D* and interfering with the desired operation of the plunger.

An improved noncontacting plunger is illustrated in Fig. 13*D*. The low-characteristic-resistance line formed by the outer surface of the line and the inner surface of the cavity wall is terminated at *C* by the impedances at *A* and *B* in series. The impedance at *B* is the input impedance of a low-loss, high-R_c, quarter-wave, shorted-end line and is extremely high. The variable impedance of the right-hand cavity reflected through a quarter-wave line to point *A* is of little importance relative to this high impedance. Consequently, the impedance at *C* approximates that of an

open circuit, and a very low impedance is produced at D. A low impedance is provided at E in a similar fashion. Plungers of this type can be made to operate satisfactorily over a frequency range of at least 2:1.[1]

A second useful form of coaxial-line cavity employs a line section short-circuited at one end and open at the other end. As indicated in Fig. 14A, the outer conductor is closed around the open end of the center conductor so as to form a completely enclosed cavity. Because of capacitance effect of the fringing electric field at the end of the center conductor, resonance occurs at frequencies for which the center conductor is somewhat less than an odd multiple of a quarter wave. The electric-field pattern for the three-quarter-wave mode is illustrated in the figure.

The cavity of Fig. 14A can be adjusted by changing the penetration of the center conductor into the cavity. Only one sliding contact is required



(A) TUNABLE RESONATOR

(B) DOUBLE QUARTER-WAVE RESONATOR

FIG. 14. Sectional views of coaxial resonators operated in odd-multiple quarter-wave modes.

for this adjustment, rather than the two contacts required by the cavity of Fig. 13A. If the spacing between the open end of the center conductor and the enclosing wall is sufficient, the fringing capacitance at the end of the line is very nearly independent of the tuning adjustment, and at constant frequency successive resonances occur at half-wave intervals of center-conductor position.

A coaxial resonator sometimes employed in uhf transmit-receive devices is illustrated in Fig. 14B. This resonator may be thought of as being formed by placing the open ends of two of the resonators of Fig. 14A together. The electric field in Fig. 14B crosses each point of the center line at right angles. Therefore a thin conducting plane could be inserted at the position of this line without changing the field pattern. Each half of the resonator may then be considered as a coaxial line less than a quarter-wave long tuned to resonance by twice the capacitance between the ends of the center conductors. A second resonant mode occurs at a slightly higher frequency. For the second mode, the voltages of the two

[1] W. H. HIGGINS, "Broad-band Noncontacting Short Circuits for Coaxial Lines," *Proc. I.R.E.*, **35** (September, 1947), 906–913.

center conductors are in phase, the conductor lengths are very little less than a quarter wave, and the field pattern is similar to that of the half-wave resonator shown in Fig. 13*B*. The first mode is used in transmit-receive devices (Art. 12, Chap. XII) because of the strong electric field provided at the gap between conductor ends.

Cavity resonators bearing considerable resemblance to quarter-wave coaxial resonators are the *re-entrant cavities* of Fig. 15. Cavities of this type are widely used in transmit-receive devices (see Art. 13, Chap. XII) and as the resonant element of klystron oscillators (see Art. 1, Chap. XI). Re-entrant cavities, especially those of the forms in Figs. 15*A* and *B*, may be thought of in one way as the result of employing the quarter-wave

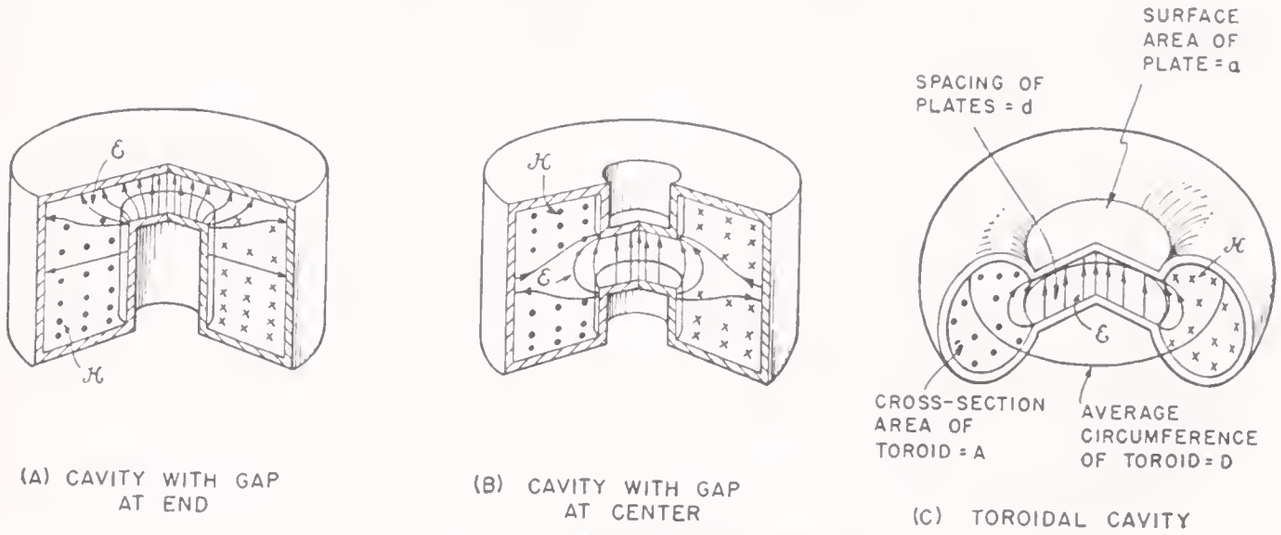

SURFACE AREA OF PLATE = a
SPACING OF PLATES = d
CROSS-SECTION AREA OF TOROID = A
AVERAGE CIRCUMFERENCE OF TOROID = D
(A) CAVITY WITH GAP AT END
(B) CAVITY WITH GAP AT CENTER
(C) TOROIDAL CAVITY

FIG. 15. Re-entrant cavities.

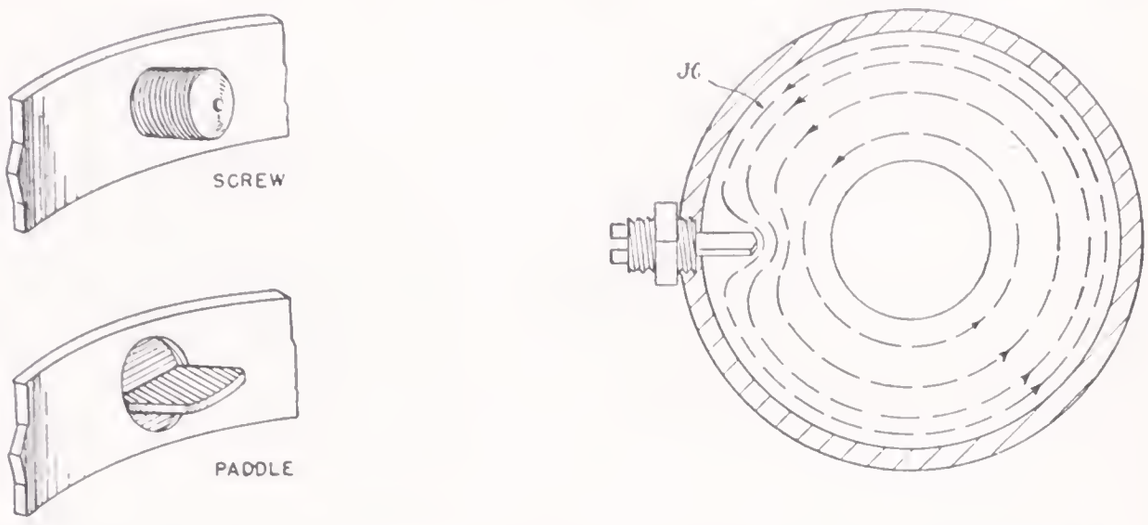
resonators of Fig. 14 at such high frequencies and with such narrow spacings between the center conductor and end wall that the line length becomes comparable to or less than the transverse dimensions. Alternatively, because the electric and magnetic fields for the most part occupy different regions in the cavity, the resonator may be thought of in terms of lumped elements. The gap occupied by the electric field may be considered a parallel-plate capacitor, and the annular magnetic field may be considered to give rise to the inductance of the conductors joining the gap electrodes.

The relation of the electric and magnetic fields to the capacitance and inductance of the cavity may be illustrated by a very approximate calculation of the resonant frequency for the cavity of Fig. 15*C*. The electric field is assumed to be uniform and wholly within the gap between center plates. Thus the capacitance C between the plates is $\varepsilon a/d$, where ε is the permittivity of the dielectric, a is the area of the plates, and d the spacing between plates. The magnetic field is assumed to be uniformly distributed in the toroidal region surrounding the plates. The magnetic-flux density set up by a current i encircled by the flux is therefore $\mu i/D$, where μ is the permeability of the dielectric and D is the path length for

the flux, assumed to be the same for all flux lines. If A is the cross-sectional area of the region occupied by the magnetic flux, the total flux is $A\mu i/D$, and the inductance that joins the capacitor plates is $\mu A/D$. The resonant frequency of the cavity is therefore

$$f_0 = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{\mu\epsilon}}\sqrt{\frac{D}{A}}\sqrt{\frac{d}{a}} \tag{24}$$

Equation 24 suggests methods of tuning the re-entrant cavity. If the separation distance d between center plates is increased, the natural frequency of the cavity is increased because of the decrease of the capaci-



(A) CUTAWAY VIEWS OF TUNING DEVICES IN RESONATOR

(B) EFFECT OF TUNING DEVICE ON MAGNETIC FIELD

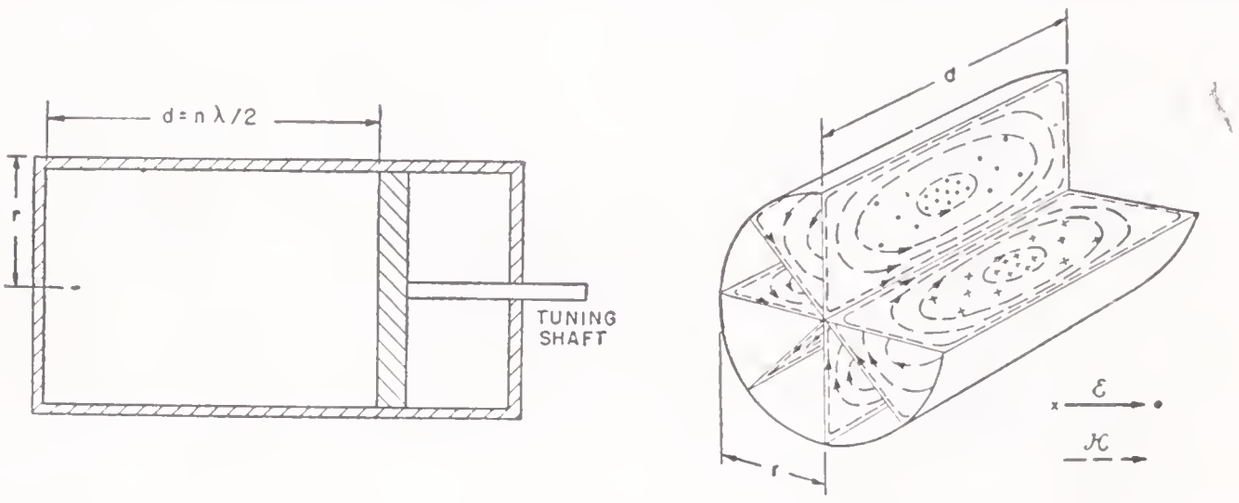
FIG. 16. Tuning screws and paddles for re-entrant cavities.

tance. Another method of increasing the frequency is to reduce the inductance by confining the magnetic field to a smaller space. Figure 16*A* shows the forms of *tuning screws* or *plugs* and *tuning paddles* used for this purpose, and Fig. 16*B* shows the effect of a screw or paddle on the magnetic field. Turning the screw so that it penetrates farther into the cavity decreases the inductance and increases the resonant frequency, and turning the paddle to a position more nearly perpendicular to the magnetic field has the same effect. These devices may be considered analogous to the brass or copper cores used to adjust the inductance of i-f coils (see Art. 21 of Chap. VI).

A third cavity type of importance is the circular-cylinder resonator formed by closing the ends of a section of circular waveguide with conducting disks. For a resonator of this form, illustrated in Fig. 17*A*, resonance occurs at frequencies for which the cavity length d is a multiple of half the guide wavelength. As with the rectangular-waveguide of Art. 4, a great variety of resonances are possible because each of the transverse-electric and transverse-magnetic waves of the circular guide

gives rise to a set of modes of oscillation. Modes are designated, as for the rectangular cavity, by adding the subscript p to the designation of the guide wave that causes the mode. The added subscript indicates the number of half guide wavelengths contained in the axial length of the cylinder. Tuning is ordinarily accomplished, as suggested in Fig. 17*A*, by moving one of the end plates.

Of particular importance are the low-loss modes designated $TE_{0,m,p}$ arising from the circular transverse-electric waves (see Art. 15, Chap. VIII). With these waves the longitudinal current components in the guide walls are eliminated, and the magnetic field intensity in the central



(A) CROSS SECTION (B) FIELDS OF $TE_{0,1,1}$ MODE

FIG. 17. Circular-cylinder resonator.

parts of the cylinder away from the conducting walls is relatively large. As a result, cavities of extremely high Q as well as guides of low attenuation are obtained. At frequencies of 10,000 Mcps, for example, values of Q at least as high as 100,000 are possible. The field configuration for the lowest frequency circular-electric mode is sketched in Fig. 17*B*. Note that the guide-wall currents, which flow along paths perpendicular to the magnetic-field lines, do not cross the junction between the cylinder wall and the movable plunger. Thus the design of a noncontacting type of plunger is simplified.

The frequencies of the circular-cylinder modes may be calculated in terms of the cutoff wavelength of circular-guide waves by means of Eq. 42 of Chap. VIII, p. 575,

$$\frac{1}{\lambda_0^2} = \frac{1}{\lambda^2} + \frac{1}{\lambda_{co}^2}$$

Because $\lambda/2$ must equal d/p, this relation takes the form

$$\left(\frac{f}{v}\right)^2 = \left(\frac{p}{2d}\right)^2 + \frac{1}{\lambda_{co}^2}$$

or

$$f = v\sqrt{\left(\frac{p}{2d}\right)^2 + \frac{1}{\lambda_{co}^2}} \tag{25}$$

where v is the speed of wave propagation in the cavity dielectric. In order for Eq. 25 to be useful, the relation of λ_{co} to the radius r of the cylinder must be known. In Table 1 values of λ_{co}/r for $TE_{0,m}$ waves are listed for values of m from 1 through 4, and an approximation formula that yields λ_{co}/r for higher values of m is given.

TABLE 1. CUTOFF WAVELENGTHS OF $TE_{0,m}$ WAVES IN CIRCULAR CYLINDERS

	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m > 4$
$\frac{\lambda_{co}}{r}$	1.64	0.89	0.62	0.47	$\frac{8}{4m+1}$

6. Impedance of One-entry Resonators. In the preceding articles resonant circuits, lines, and cavities have been considered as isolated oscillating systems. To be useful, these resonators must be connected to other apparatus. The remainder of this chapter is therefore devoted to a discussion of the effect of resonators upon associated circuit components. This article considers resonant systems operating at a frequency sufficiently low so that a definite pair of input terminals can be established. It is assumed that the only connection between the resonator and the rest of the circuit is at these terminals. Such systems may be classified as *one-entry* resonators.

An important property of a one-entry resonator is its *input impedance* at the point of entry. It may be shown that the input impedance of any lossless resonator is a pure reactance which varies with frequency in the general manner shown in Fig. 18. At certain critical frequencies, such as f_1, f_2, and f_3, the input reactance becomes infinite, and at other frequencies, such as f_a, f_b, and f_c, the reactance is zero. Frequencies corresponding to infinite reactance are referred to as *poles*, and frequencies corresponding to zero reactance are called *zeros*. A network composed entirely of lumped-constant reactive elements always has a finite number of poles and zeros, but if lossless transmission lines are included, the number of poles and zeros becomes infinite.

The input-reactance curves of one-entry resonators have several important properties which may be summarized as follows:[1] (1) The zeros and poles alternate along the frequency axis, (2) the slope of the curve is

[1] See, for example, E. A. GUILLEMIN, *Communication Networks*, Vol. II, 184–187 (John Wiley & Sons, Inc., New York, 1935).

always positive, and (3) any two reactance curves having the same poles and zeros (including those at zero and infinity) are identical at all frequencies except for a scale factor of reactance. Because of these properties, the pattern of poles and zeros is nearly a complete description of the reactance curve of a lossless resonator.

The poles and zeros of the input-reactance curve of a one-entry lossless network are closely related to the natural frequencies of the network. Poles correspond to modes of oscillation for which an input voltage but no input current is present and thus represent natural frequencies deter-

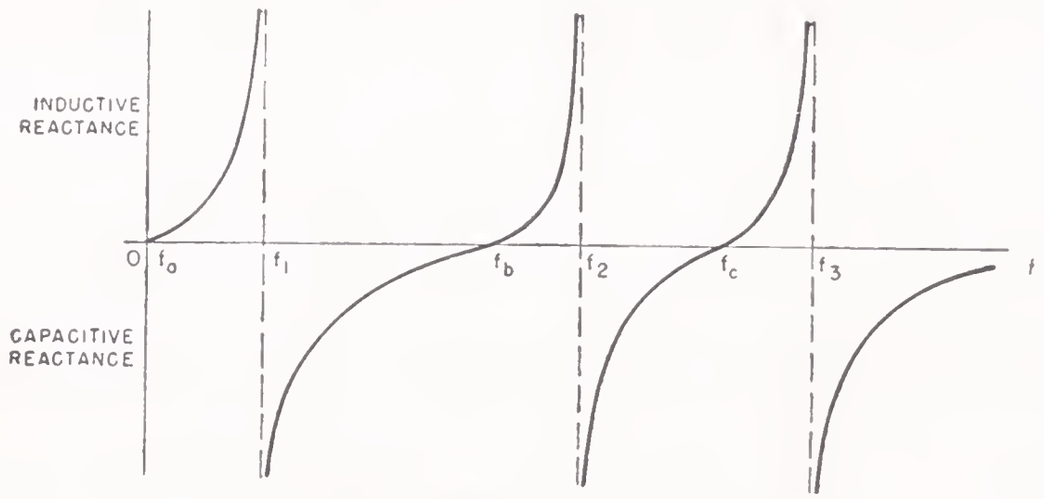


FIG. 18. Input reactance of a lossless network.

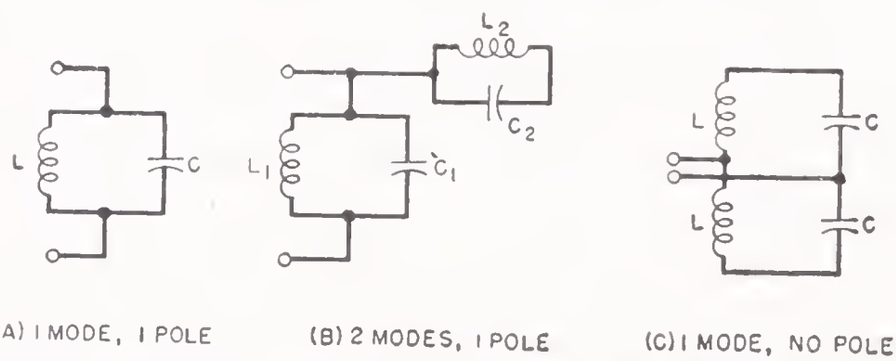


FIG. 19. Number of open-circuit modes and poles for several networks.

mined with the input terminals open-circuited. Similarly, at zeros, an oscillating input current is obtained with zero voltage. Zeros therefore represent natural frequencies determined with the input terminals short-circuited.

Although each pole corresponds to a natural frequency with input terminals open-circuited, every natural frequency of a network does not necessarily produce a pole in the reactance curve. A pole occurs only if oscillations in the mode actually cause a voltage across the input terminals. The simple *L-C* resonant circuit in Fig. 19, for example, has one natural mode with the input terminals open. Because oscillations in the mode cause a voltage between input terminals, the input-reactance curve has a pole corresponding to the natural frequency. In contrast, the

middle network has two open-circuit natural modes but only one pole, because oscillations in the L_2, C_2 branch are not coupled to the input terminals. The reactance curve of the network at the right has no poles because, at resonance, no voltage appears at the input terminals. In a similar way, natural modes with the input terminals short-circuited yield

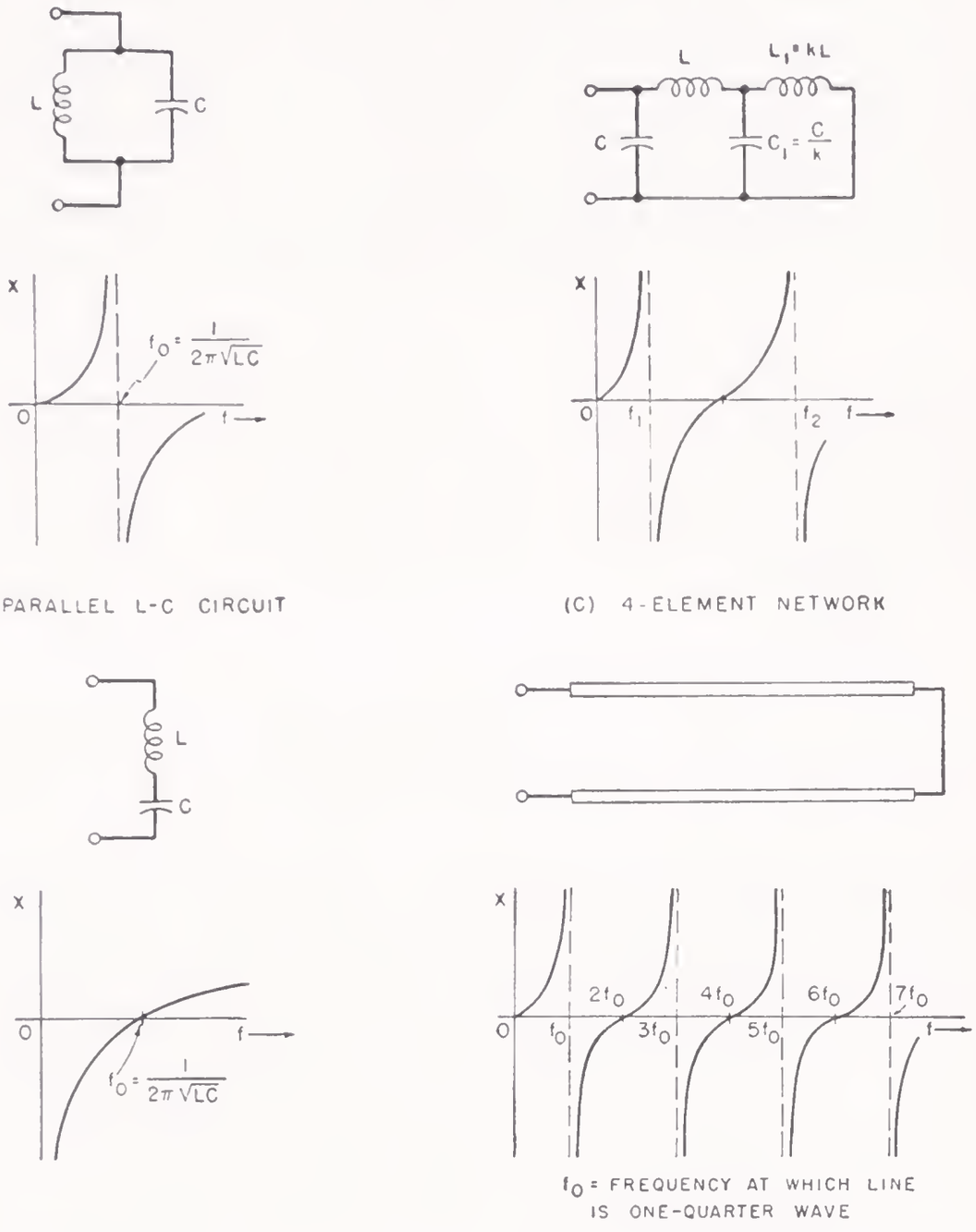


FIG. 20. Reactance curves of one-entry networks.

zeros in the reactance curve only if the oscillations cause current flow at the input terminals.

Several one-entry resonators and their reactance curves are drawn in Fig. 20. The two resonators at the left are simple L-C resonant circuits and have the familiar curves shown. The four-element network is the network of Fig. 2 with a pair of input terminals added. The two poles in the reactance curve correspond to natural frequencies determined with the

input terminals open (see Art. 1). The zero between the poles, determined with the input terminals short-circuited, represents the natural frequency of the parallel combination of L, L_1, and C_1. Zeros also occur at zero frequency (where L and L_1 are short circuits) and at infinite frequency (where C and C_1 are short circuits). The input-reactance curve of the shorted-end line is discussed in Art. 3 of Chap. VII. The curve has poles and zeros equally spaced along the entire frequency axis.

When losses in a network are taken into account, the input impedance is no longer a pure reactance. However, *small* losses have negligible effect upon input impedance except near poles and zeros. In the region of a pole, the reactive component of current becomes very small and the small in-phase current required to supply losses becomes important.

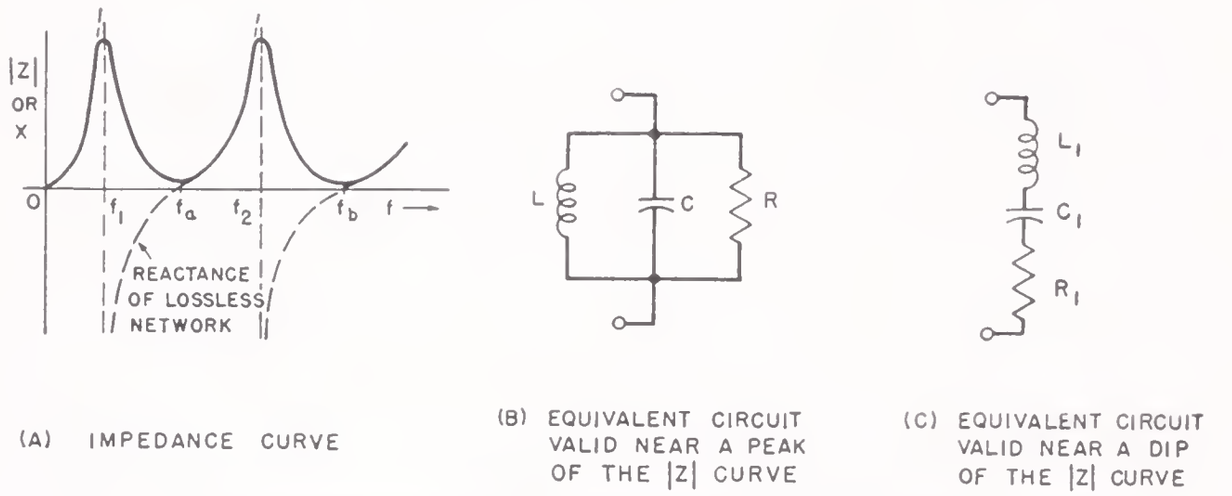


FIG. 21. Impedance curve and equivalent circuits for a one-entry resonator.

Similarly, near each zero the reactive component of voltage becomes small and the in-phase voltage needed to supply losses becomes important. Thus the input impedance at a pole, instead of being infinite, is a large resistance, and at a zero the impedance becomes a small resistance rather than zero. At all other frequencies the quadrature component of current or voltage is large compared with the corresponding in-phase component and the input impedance is very nearly a pure reactance.

For a one-entry resonator with small losses the magnitude of the input impedance varies with frequency in the general way illustrated by the curve of $|Z|$ in Fig. 21*A*. As indicated in the figure, the $|Z|$ curve may be derived from the reactance curve that applies to the resonator if losses are ignored. Except in small frequency bands surrounding peaks and dips, the impedance of the actual resonator is essentially equal to the reactance of the lossless resonator and is inductive at frequencies above a dip and below a peak and capacitive at frequencies above a peak and below a dip. Near peaks and dips the impedance passes through complex values and becomes a resistance at almost exactly the frequencies of maximum or minimum $|Z|$. The smaller the loss (or the higher the Q) of each natural

mode, the greater is the resistance at the peak or the more nearly zero is the resistance at the dip.

Often only a small frequency band near one of the peaks or dips of an impedance curve is of interest. The impedance curve of Fig. 21*A* suggests a *parallel R-L-C circuit* (Fig. 21*B*) as an equivalent circuit for the resonator in the region near a *peak* and a *series R-L-C circuit* (Fig. 21*C*) as an equivalent circuit near a *dip*. It may in fact be demonstrated that these equivalent circuits are always valid for a one-entry resonator, provided the critical frequency of interest is well separated from adjacent critical frequencies. The values of R, L, and C must, of course, be chosen so that the impedance curve of the equivalent circuit is a duplicate of the impedance curve of the resonator in the frequency band of interest. The properties of the $|Z|$ curve needed to make the calculation are the height, sharpness, and frequency of the peak or dip being represented.

Suppose that a peak of height $|Z|_{\max}$ is located at an angular frequency $\omega_0 = 2\pi f_0$ and that the bandwidth between the frequencies for which $|Z|$ is $|Z|_{\max}/\sqrt{2}$ is Δf. The equivalent circuit valid near the peak takes the form of Fig. 21*B*. The values of the elements may be determined from the relations developed in Art. 6 of Chap. VI for parallel *R-L-C* circuits; that is,

$$R = |Z|_{\max}, \qquad \sqrt{LC} = \frac{1}{\omega_0}, \qquad \sqrt{\frac{L}{C}} = \frac{R}{Q} \tag{26}$$

where $1/Q = \Delta f/f_0$ is the per-unit bandwidth of the peak being represented. The parameter $\sqrt{L/C} = \omega_0 L = 1/\omega_0 C$ is the reactance of L or C at resonance and is therefore a convenient measure of the impedance level at resonance. Observe that, for a fixed value of Q, the impedance R at resonance is directly proportional to $\sqrt{L/C}$.

The values of R, L, and C in the equivalent parallel circuit may also be calculated from the parameters of the natural mode of resonance that causes the peak of impedance. These parameters are the natural frequency of the mode f_0, the quality factor Q, and the stored energy W for a specified peak voltage E_m at the open-circuited input terminals. These parameters are determined from the physical form of the resonant system by the methods given in Arts. 1 through 5.

For the equivalent circuit to have the same input impedance as the actual resonator in a frequency band near an open-circuit resonance, not only must the equivalent circuit have a natural frequency equal to that of the resonator but also at resonance it must store the same amount of energy and dissipate the same amount of power for a specified terminal voltage. That the power losses must be equal is evident in order that the resonator and its equivalent may have the same resistance at resonance. That the stored energies must also be equal is not so apparent but may

be proved.[1] The equal-energy condition is necessary in order for the equivalent circuit to duplicate the sharpness of the $|Z|$ curve of the resonator.

The need for equal stored energies may be explained qualitatively in the following manner: When a resonator is driven at one of its natural frequencies, a constant amount of energy oscillates between the electric and magnetic fields. The source supplies only the power consumed by losses. At other frequencies the peak energies in the electric and magnetic fields are unequal, and some energy from the source is stored during one part of the cycle and returned to the source during another part of the cycle. The storing and returning of energy require a reactive component of current from the source and account for the decrease in impedance at off-resonance frequencies. The more energy stored in the oscillation, the greater is the unbalance of peak energies in the electric and magnetic fields for a given frequency change from f_0 and the sharper is the decrease in input impedance. A large stored energy corresponds to a low impedance level for the reactive part of the equivalent circuit. The low value of $\sqrt{L/C}$, together with a high shunt resistance R, provides a high Q and hence a very sharp resonant peak in the $|Z|$ curve of the equivalent circuit.

The values of R, L, and C in the equivalent circuit of a natural mode are related to the mode parameters ω_0, Q, and W (for specified E_m) in the following manner: The energy stored in the equivalent circuit, $\frac{1}{2}CE_m^2$, must equal W, the energy stored in the actual circuit for the same voltage. Thus $W = \frac{1}{2}CE_m^2$, and the capacitance of the equivalent circuit has the value

$$C = \frac{2W}{E_m^2} \tag{27}$$

In order for the equivalent circuit to resonate at an angular frequency ω_0, LC must equal $1/\omega_0^2$. Hence

$$L = \frac{1}{\omega_0^2 C} \tag{28}$$

Since both the equivalent circuit and actual resonator must have the same Q, the shunt resistance R of the equivalent circuit is

$$R = Q\sqrt{\frac{L}{C}} \tag{29}$$

where Q is the quality factor of the natural mode being represented.

If the frequency band near a dip in the $|Z|$ curve is of interest, the resonator may be represented by a series resonant circuit (Fig. 21*C*).

[1] E. A. GUILLEMIN, *Communication Networks*, Vol. II, 226–229 (John Wiley & Sons, Inc., New York, 1935).

The values of R_1, L_1, and C_1 are determined from the relations

$$R_1 = |Z|_{\min}, \qquad \sqrt{L_1 C_1} = \frac{1}{\omega_0}, \qquad \sqrt{\frac{L_1}{C_1}} = RQ \tag{30}$$

where $\omega_0 = 2\pi f_0$ is the angular frequency at the dip, $1/Q = \Delta f/f_0$ is the per-unit bandwidth, and Δf is the frequency band between points where $|Z|$ is equal to $\sqrt{2}\,|Z|_{\min}$. Alternatively, R_1, L_1, and C_1 may be calculated from mode parameters obtained with the input terminals short-circuited.

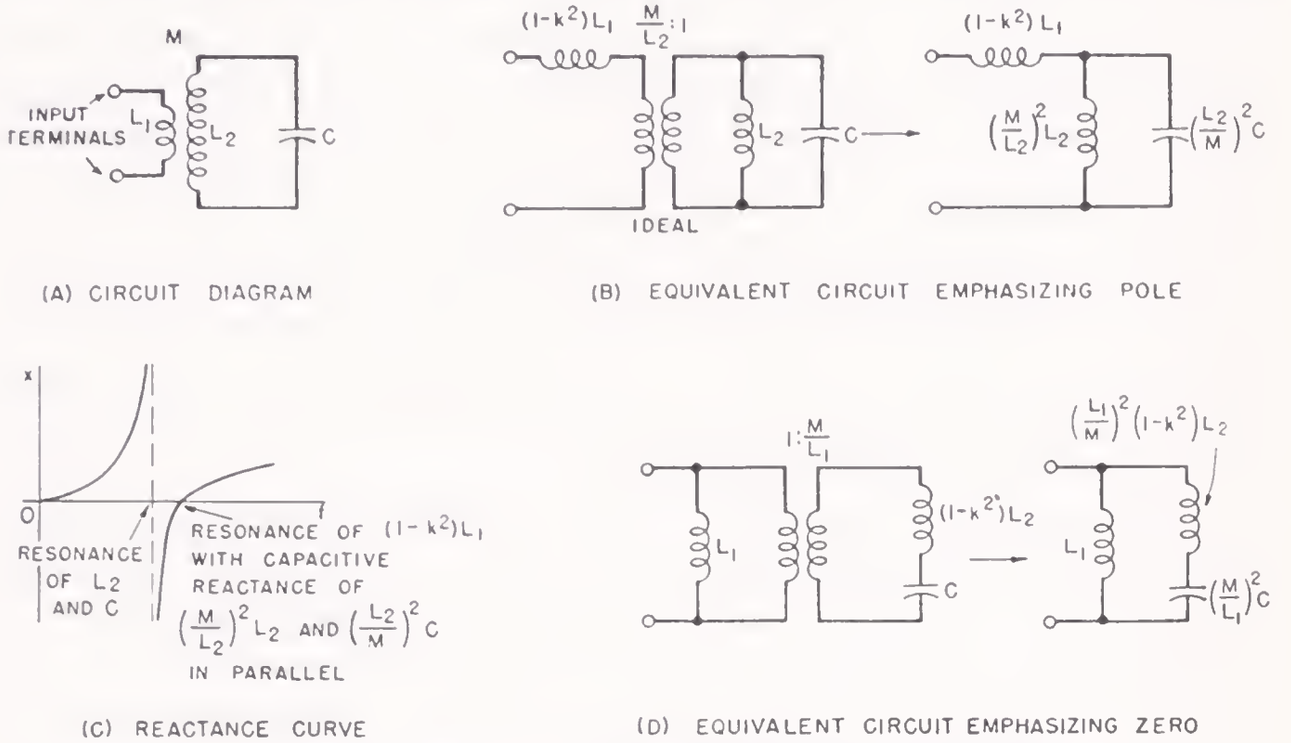


FIG. 22. Lossless resonator with transformer coupling.

Thus, if I_m is the peak current at the input terminals and W is the energy stored for the mode of interest,

$$L_1 = \frac{2W}{I_m^2}, \qquad C_1 = \frac{1}{\omega_0 L}, \qquad R_1 = \frac{\sqrt{L/C}}{Q} \tag{31}$$

7. Coupling to One-entry Resonant Networks. Although one-entry resonant networks may be connected directly to other apparatus, direct connection is often undesirable because of the extremely high impedance of the resonator near its poles. A suitable coupling device, in theory, is an ideal transformer having a turns ratio chosen to make the impedance at its input terminals a moderately high resistance at resonance and almost zero at slightly different frequencies. Ideal transformers do not exist, of course, but several practical coupling devices are available to reduce the impedance level of the resonator in very nearly the same way as would an ideal transformer.

One useful device is an actual transformer having a large turns ratio. As indicated in Fig. 22*A*, the secondary winding has the larger number of

turns and may serve as the inductance element of the resonator. An equivalent circuit of the resonator is drawn in Fig. 22*B* in terms of an ideal transformer (see Art. 20, Chap. II). The turns ratio M/L_2:1 of the ideal transformer is much less than unity, and the desired reduction in resonator impedance is achieved.

Observe in Fig. 22*B* that, if the coefficient of coupling k is unity, inductance $(1 - k^2)L_1$ becomes zero, and the actual transformer has only the impedance-changing effect of the ideal transformer. If k is less than unity, however, $(1 - k^2)L_1$ is effectively in series with the parallel combination of L_2 and C referred to the primary of the ideal transformer. The reactance curve of the resonator (Fig. 22*C*) then contains a *pole* at the frequency for which L_2 and C are in parallel resonance and a *zero* at a higher frequency for which $(1 - k^2)L_1$ is in series resonance with the equivalent capacitance of the parallel combination of $(M/L_2)^2L_2$ and $(L_2/M)^2C$. If the reactance of $(1 - k^2)L_1$ is large relative to the resonant impedance of the tuned circuit, the zero lies close to the pole on the frequency axis. Thus two closely spaced critical frequencies are obtained if the coupling coefficient k is small or if a small ratio M/L_2 is employed to obtain a large ratio of transformation.

The series reactive element $(1 - k^2)L_1$ is called the *coupling reactance* of the network. Coupling reactance occurs not only in transformer coupling but also in many other coupling devices to be considered. A coupling device that provides a large reduction of impedance level of a parallel-resonant system almost always introduces an important coupling reactance. Because of the coupling reactance, a short-circuit natural mode is obtained at a frequency close to that of the open-circuit mode, and therefore a closely spaced pole-zero pair appears in the reactance curve.

The equivalent circuit of Fig. 22*B* emphasizes the pole of the pole-zero pair inasmuch as the parallel circuit producing the pole appears directly. Equivalent circuits of the form of Fig. 22*B* are especially desirable when frequencies in the region of the pole are of interest. If the frequency band near the zero is more important, the alternative circuit shown in Fig. 22*D* becomes useful. This circuit, which is obtained by using a different equivalent circuit (see Art. 20, Chap. II) for the transformer of Fig. 22*A*, emphasizes the zero or frequency at which $(1 - k^2)L_2$ and C are in series resonance. The pole in the reactance curve is still accounted for because at a frequency below the zero the series circuit referred to the transformer primary is effectively a capacitance that resonates in parallel with L_1. Inductance L_1 may therefore be regarded as a coupling *susceptance*. It may be concluded that a resonator may be represented in the region of a pole-zero pair in its reactance curve either as a *parallel resonant circuit in series with a coupling reactance* or as a *series resonant circuit shunted by a coupling susceptance*.

If only a very narrow band of frequencies is of interest near a pole of a *lossless* resonator, coupling reactance may be ignored because the reactance of the tuned circuit increases without limit as the resonant frequency is approached. Similarly, coupling susceptance may be disregarded in a series-circuit representation of a short-circuit mode if the system is lossless and if only a small frequency band near the zero is important. In a system with losses, however, the impedance (or admittance) at a pole (or zero) is limited by the loss, and the coupling reactance (or susceptance) may be important even at resonance. Therefore, a coupling reactance may be needed in an equivalent circuit in order completely to represent

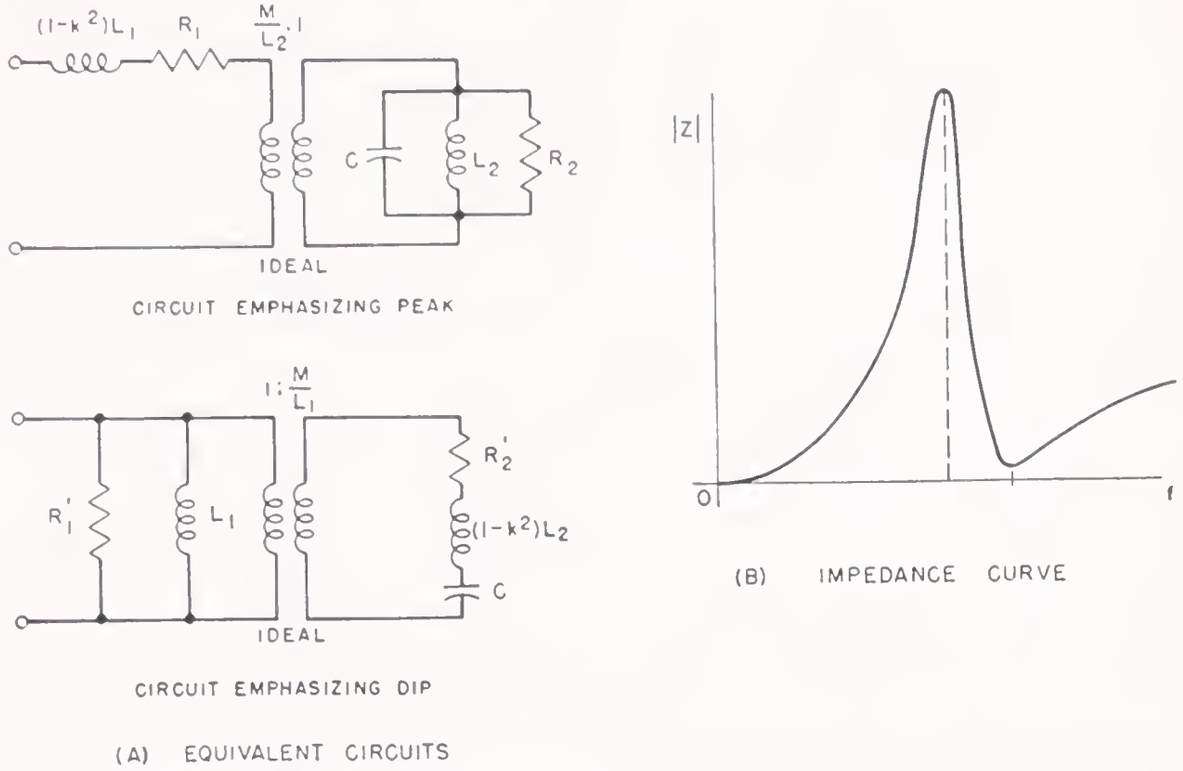


FIG. 23. Equivalent circuits and impedance curve for resonator of Fig. 22*A* with losses included.

a practical resonator, regardless of the smallness of the frequency band being considered.

When losses are included in the resonator of Fig. 22*A*, the equivalent circuits take the forms shown in Fig. 23*A* and the magnitude of the input impedance varies with frequency as indicated in Fig. 23*B*. Resistance R_2 in the circuit emphasizing the peak is determined from the Q of the open-circuit mode by the methods described in Art. 6, and resistance R_2' in the second circuit is similarly determined from the Q of the short-circuit mode. Resistor R_1 is chosen to make the first circuit specify the correct resistance at the dip of the $|Z|$ curve, and resistor R_1' adjusts the resistance at the peak specified by the second circuit.

Connecting a resonator to a circuit through a transformer introduces an inductive coupling reactance or susceptance and causes the dip in the $|Z|$ curve to occur at a higher frequency than the peak. Other coupling

methods may introduce a capacitive coupling reactance or susceptance and place the dip below the peak in the $|Z|$ curve. For example, the capacitive branch of a parallel L-C current may be split into two capacitors in series, as in Fig. 24A, and the resulting capacitive voltage provides a means of reducing the impedance level of the resonant circuit. To obtain the required reduction in impedance, C_0 in the figure must be made much greater than C.

The circuit of Fig. 24A emphasizes the dip in the $|Z|$ curve of Fig. 24B, the dip being caused by series resonance of L and C. Because of the coupling susceptance of C_0, a peak in the $|Z|$ curve occurs at a higher fre-

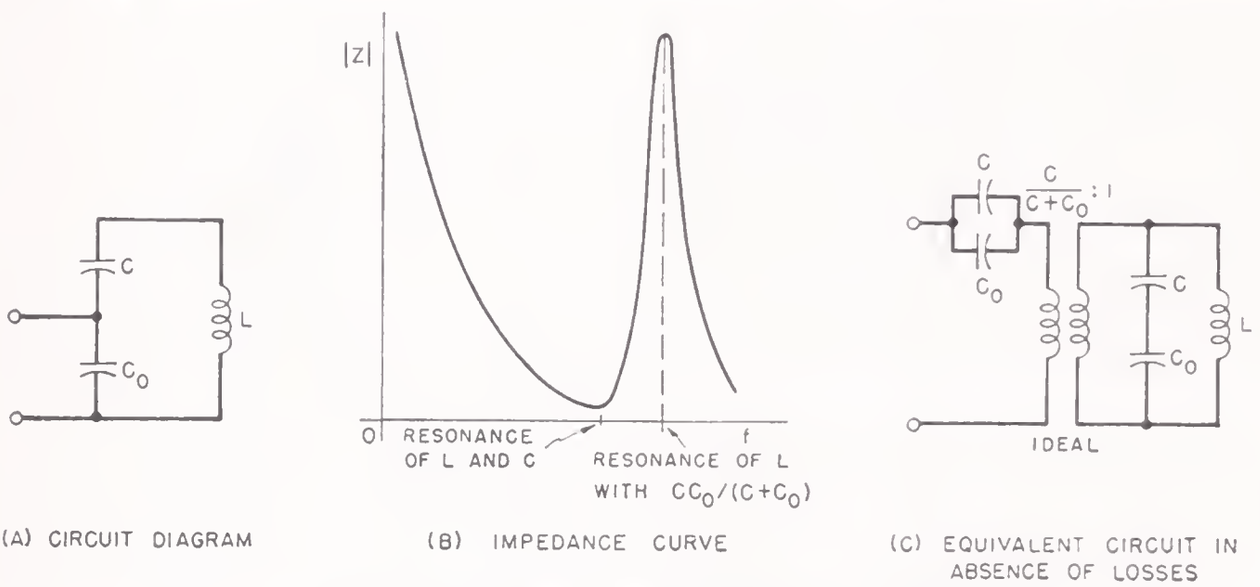


FIG. 24. Resonator with capacitive coupling.

quency. The peak is made evident by the equivalent circuit of Fig. 24C, obtained from Fig. 24A by replacing the capacitance voltage divider by an ideal transformer with shunt and series capacitors (see Art. 20, Chap. II). If C_0 is made large so that reduction of impedance level provided by the transformer of Fig. 24C is great, the two critical frequencies are very close together and the coupling reactance or susceptance is important.

The above examples of coupling devices illustrate two ways in which coupling reactance or susceptance may be introduced into the equivalent circuit of a resonator. In general, whenever the natural frequency of a short-circuit resonant mode lies close to the frequency of an open-circuit mode of interest, coupling reactance in series with a parallel R-L-C circuit is required for accurate representation of the resonator in a range near the open-circuit natural frequency.

The value of the coupling reactance can be calculated, but an exact determination is rather lengthy. It is necessary first to obtain an equivalent circuit that, ignoring losses, accounts for all poles and zeros of the reactance curve. One such circuit,[1] consisting of a series chain of parallel

[1] E. A. GUILLEMIN, *Communication Networks*, Vol. II, 184–194 (John Wiley & Sons, Inc., New York, 1935).

L-C circuits, is represented in Fig. 25. A series capacitance is added to this circuit if the reactance curve has a pole at zero frequency, and a series inductor is added if the curve has a pole at infinite frequency. Each parallel circuit represents an open-circuit mode of oscillation and gives rise to a pole in the reactance curve. In the frequency range of interest, all but one of the parallel circuits is a reactance of essentially constant value, and the algebraic sum of the essentially constant reactances is the coupling reactance.

Ordinarily the coupling reactance is only a small part of the total impedance of a resonator. Therefore, calculation of its exact value by a lengthy procedure requiring determination of the entire circuit of Fig. 25 is unnecessary; a simpler calculation yielding an approximate value for the coupling reactance may be used instead. If the resonator has a d-c path between its input terminals, the coupling reactance for low-frequency modes is usually of the same order of magnitude as the reactance of the

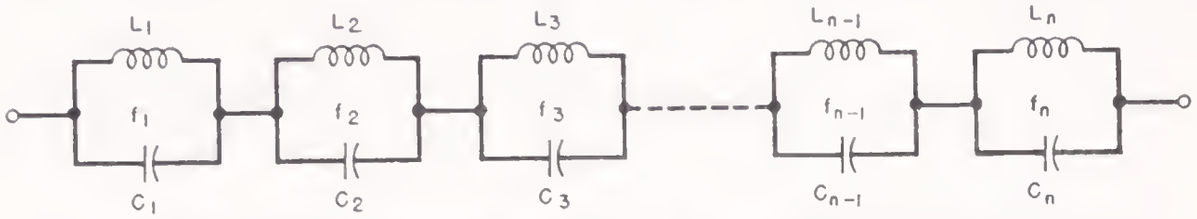


FIG. 25. Complete equivalent circuit of a lossless one-entry resonator.

inductance measured at very low frequencies between the resonator terminals. In Fig. 25, for example, if frequencies near f_2 (the resonant frequency of L_2 and C_2) are of interest, and if f_1 is less than f_2, f_2 is less than f_3, and so on, then the coupling reactance is the series combination of the capacitive reactance of the L_1, C_1 circuit (at frequency f_2) and the inductive reactance of the circuits resonant at frequencies above f_2. This coupling reactance is not greatly different from the reactance of the low-frequency inductance of the resonator $L_1 + L_2 + \cdots + L_n$, provided L_1 and L_2 form only a small part of the total low-frequency inductance. If the lowest frequency mode is of interest, the value of L_1 is known, and it is often worth while to subtract L_1 from the low-frequency inductance in order to obtain a more accurate value of coupling inductance.

If a d-c path does not exist between resonator terminals, a capacitor C_0 appears in series with the tuned circuits of Fig. 25, and C_0 is the capacitance measured between resonator terminals at very low frequencies. Often the reactance of C_0 at the resonant frequency may be used as an approximate value for coupling reactance, although the actual value of the coupling reactance includes the reactances of the L-C circuits other than the one resonating in the frequency range of interest.

8. Equivalent Circuits for Transmission-line Resonators. The principles developed in Arts. 6 and 7 may be applied to transmission-line

resonators in order to obtain equivalent circuits valid near poles or zeros of the input-reactance curves. Although low-loss parallel-wire lines are assumed in the figures of this article, the equivalent circuits to be discussed apply equally well to coaxial lines.

Figure 26*A* represents a portion of transmission line of length d, either open-circuited or short-circuited at the far end. In absence of losses, the input reactance varies with frequency in the manner shown in Fig. 26*B* if the far end is open or as indicated in Fig. 26*C* if the far end is short-circuited. The lowest frequency pole of the shorted-end line and the

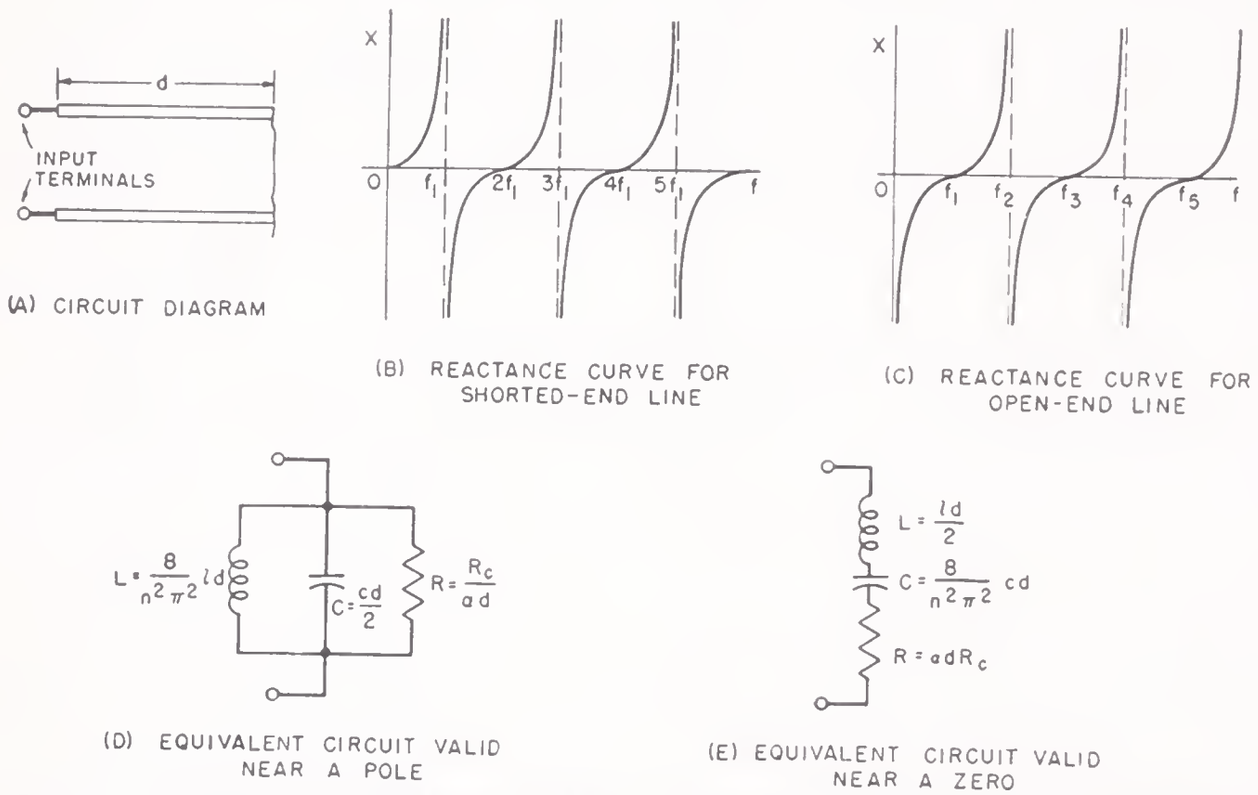


FIG. 26. Reactance curves and equivalent circuits for a transmission-line resonator.

lowest frequency zero of the open-end line occur at a frequency f_1 for which the line length d is a quarter wave. As explained in Art. 3,

$$f_1 = \frac{1}{4d\sqrt{lc}} \tag{32}$$

where l and c denote the inductance and capacitance per-unit length of the line.

First consider the equivalent parallel circuit of Fig. 26*D*, which represents the line near a pole of its reactance curve. Let the frequency of the pole be $f_0 = nf_1$, where n, the number of quarter waves on the line at f_0, is an odd integer for a shorted-end line and an even integer for an open-end line. If the peak voltage at the pole is $E_{m_{max}}$, the energy stored on the line is

$$W = \tfrac{1}{4}cdE^2_{m_{max}}$$

and since the capacitance C in the equivalent parallel circuit must store the same amount of energy at the same voltage, C must equal $2W/E^2_{m_{max}}$, or

$$C = 2\left(\frac{\frac{1}{4}cdE^2_{m_{max}}}{E^2_{m_{max}}}\right) = \frac{1}{2}cd \tag{33}$$

Thus the equivalent-circuit capacitance is equal to one-half the total capacitance of the line.

The inductance L in the equivalent circuit must be chosen to provide the correct resonant frequency. Thus L is equal to $1/\omega_0^2 C$, or by Eqs. 31 and 32

$$L = \frac{1}{\omega_0^2 C} = \frac{1}{(2\pi n f_1)^2 C} = \left(\frac{4d\sqrt{lc}}{2\pi n}\right)^2 \left(\frac{2}{cd}\right) = \frac{8}{n^2\pi^2} ld \tag{34}$$

To represent the effect of small line losses, the equivalent shunt resistance R must have the value $Q\sqrt{L/C}$, where (by Eq. 12, p. 624) Q equals $\pi/\lambda\alpha$, the quality factor of the line. Thus

$$R = Q\sqrt{\frac{L}{C}} = \frac{\pi}{\lambda\alpha}\sqrt{\left(\frac{8ld}{n^2\pi^2}\right)\left(\frac{2}{cd}\right)} = \frac{4}{n\lambda\alpha}\sqrt{\frac{l}{c}}$$

The factor $\sqrt{l/c}$ may be replaced by R_c, the characteristic resistance of the line, and $n\lambda/4$ may be replaced by d, with the following result:

$$R = \frac{R_c}{\alpha d} \tag{35}$$

Thus the shunt resistance in the equivalent circuit is equal to the characteristic resistance of the line divided by the total attenuation of the line.

Observe that C in Fig. 26D is independent of n and therefore has the same value at all poles. In contrast, L is inversely proportional to n^2 and hence decreases as the order of the pole increases. The impedance-level parameter $\sqrt{L/C}$ therefore varies in inverse proportion to frequency, whereas R is inversely proportional to $\sqrt{f}$ (because α is proportional to $\sqrt{f}$). The quality factor Q is therefore proportional to $\sqrt{f}$.

Coupling reactance is omitted in Fig. 26D because the rather wide separation of poles and zeros suggests that coupling reactance should be negligible. By means of a rather lengthy calculation it may be shown that for either open-end or shorted-end lines the coupling reactance at each pole is inductive and equal to $\frac{1}{4}\sqrt{L/C}$. For all reasonable values of Q, therefore, the equivalent circuit has an impedance at frequencies near resonance which is at least several hundred times the coupling reactance.

That $\frac{1}{4}\sqrt{L/C}$ represents the coupling reactance may be verified approximately for a quarter-wave shorted-end line in the following man-

ner: The low-frequency inductance of the line is ld, and from Eq. 34 the inductance of the equivalent circuit for $n = 1$ is $8ld/\pi^2$. As explained in Art. 7, an approximate value for the coupling inductance is therefore $ld - 8ld/\pi^2$ or $0.19ld$. At quarter-wave resonance the ratio of the coupling reactance to the equivalent-circuit inductive reactance is therefore $0.19ld/(8ld/\pi^2)$, which is equal to 0.234 or approximately 0.25.

A series R-L-C equivalent circuit valid near a zero of the transmission-line reactance curve may be constructed from parameters of the short-circuit modes. In terms of $I_{m_{\max}}$, the peak value of the maximum current along the line, the stored energy W is $\frac{1}{4}ldI^2_{m\max}$. An inductance L that stores the same amount of energy for the same peak current has the value

$$L = \frac{2W}{I^2_{m_{\max}}} = 2\left(\frac{\frac{1}{4}ldI^2_{m_{\max}}}{I^2_{m_{\max}}}\right) = \frac{1}{2}ld \tag{36}$$

or one-half the total inductance of the line. The series capacitance C required to resonate with L at the frequency of the zero ($f_0 = nf_1$) is $C = 1/\omega_0^2 L$, and by Eq. 32,

$$C = \frac{1}{\omega_0^2 L} = \frac{1}{(2\pi n f_1)^2 L} = \left(\frac{2d\sqrt{lc}}{n\pi}\right)^2\left(\frac{2}{ld}\right) = \frac{8}{n^2\pi^2}cd \tag{37}$$

The series resistance R that produces a Q for the equivalent circuit equal to that of the line has the value

$$R = \frac{1}{Q}\sqrt{\frac{L}{C}} = \frac{\lambda\alpha}{\pi}\sqrt{\left(\frac{ld}{2}\right)\left(\frac{n^2\pi^2}{8cd}\right)} = \frac{n\lambda\alpha}{4}\sqrt{\frac{l}{c}} = \alpha d R_c \tag{38}$$

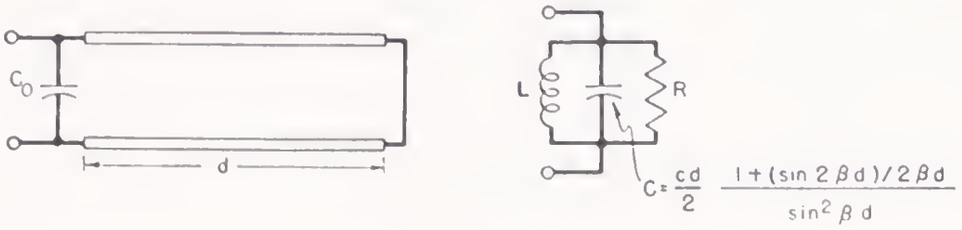
That is, R is equal to the characteristic resistance of the line times the attenuation of the line.

The series equivalent circuit is drawn in Fig. 26E. In this circuit L has the same value at all zeros, whereas C becomes smaller as the frequency of the zero increases. The quality factor at resonance also increases with the frequency of the zero because R, being directly proportional to α, varies as $\sqrt{f}$ whereas $\sqrt{L/C}$ is directly proportional to f.

Another transmission-line resonator of interest is the shorted-end line tuned by a capacitor connected across the input terminals. A circuit diagram and an equivalent circuit applicable to the lowest frequency open-circuit mode are drawn in Fig. 27. The value for C indicated in Fig. 27B may be determined from the stored energy of the natural-mode oscillation. By Eq. 10, p. 623, this energy is

$$W = \frac{1}{2}ldI^2_{m_{\max}}\left(\frac{1 + (\sin 2\beta d)/2\beta d}{2}\right)$$

where $I_{m_{max}}$ is the peak value of the current at the short-circuited end of the line and βd is the electrical length of the line at the frequency of oscillation. According to Eqs. 21 and 22 of Chap. VII, $I_{m_{max}}$ has the value $E_{m_{max}}/R_c$, where $E_{m_{max}}$ is the peak value of the voltage required at the open end of a quarter-wave line to produce the current $I_{m_{max}}$ in a short circuit at the other end of the line. In terms of E_m, the peak voltage at the input



(A) CIRCUIT DIAGRAM

(B) EQUIVALENT CIRCUIT VALID NEAR LOWEST FREQUENCY PEAK OF |Z| CURVE

FIG. 27. Resonator formed by a shorted-end line and a shunt capacitor.

terminals of the line of Fig. 27*A*, $E_{m_{max}} = E_m/\sin \beta d$. The stored energy may therefore be expressed as

$$W = \frac{1}{2} ld \left(\frac{E_m}{\sin \beta d} \sqrt{\frac{c}{l}} \right)^2 \left(\frac{1 + (\sin 2\beta d)/2\beta d}{2} \right) = \frac{1}{4} cd E_m^2 \left(\frac{1 + (\sin 2\beta d)/2\beta d}{\sin^2 \beta d} \right)$$

An equivalent parallel *R-L-C* circuit that stores the same amount of energy at a peak voltage E_m has a capacitance $C = 2W/E_m^2$, and therefore

$$C = \frac{1}{2} cd \left(\frac{1 + (\sin 2\beta d)/2\beta d}{\sin^2 \beta d} \right) \tag{39}$$

If desired, the inductance required to resonate with C and the shunt resistance needed to produce the correct Q may also be calculated, the general method being the same as that employed in the preceding examples.

Note that the capacitance in the equivalent circuit is not equal to the tuning capacitance C_0. In fact, as βd approaches $\pi/2$, C tends toward $\frac{1}{2}cd$, and the equivalent circuit approaches that of Fig. 26*D*, as may be expected. On the other hand, if βd is very nearly zero,

$$\frac{1 + (\sin 2\beta d)/2\beta d}{\sin^2 \beta d} \approx \frac{2}{(\beta d)^2}$$

and C then has the approximate value $cd/(\beta d)^2$. The value of the tuning capacitance C_0 may be computed, since its reactance must equal the inductive reactance of the line at resonance. Therefore,

$$C_0 = \frac{1}{\omega_0 R_c \tan \beta d} = \frac{\sqrt{lc}\, d}{\beta d} \sqrt{\frac{c}{l}} \frac{\cos \beta d}{\sin \beta d}$$

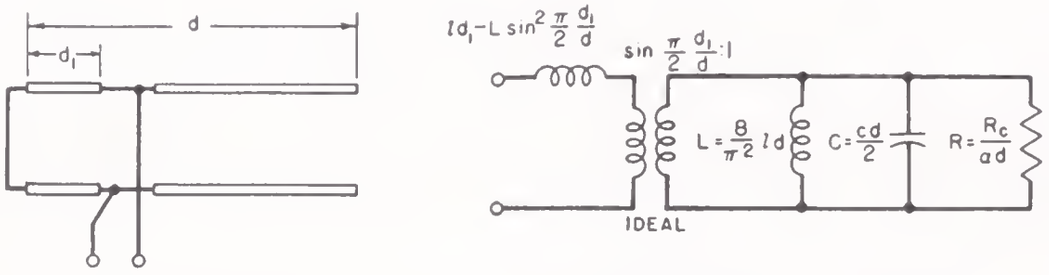
and

$$C_0 = cd \frac{\cos \beta d}{\beta d \sin \beta d}$$

For extremely small values of βd, C_0 is approximately equal to $cd/(\beta d)^2$, and therefore C approaches equality with C_0 when βd approaches zero.

9. Coupling to Transmission-line Resonators. Just as a parallel R-L-C circuit may have an impedance too high for many applications, so also a transmission-line resonator may have too high an impedance when operated near a peak of its $|Z|$ curve. It is often necessary, therefore, to couple transmission-line resonators to other circuits in a way that reduces the impedance level to the desired value. Several coupling methods are described in this article.

Figure 28A shows a line of length d short-circuited at one end with input terminals located a short distance d_1 from the short circuit. As



(A) CIRCUIT DIAGRAM

(B) EQUIVALENT CIRCUIT VALID NEAR LOWEST FREQUENCY PEAK OF |Z| CURVE

FIG. 28. Coupling to a quarter-wave line by means of taps on the conductors.

indicated in Fig. 28B, the equivalent circuit valid near the lowest frequency peak of the $|Z|$ curve is almost the same as the one obtained for quarter-wave resonance when connections are made at the end of the line—the values of R, L, and C in Fig. 28B are the same as the values for $n = 1$ in Fig. 26D. Two important differences may be noted, however: (1) The impedance level is lower, and (2) the coupling reactance may be appreciable. For a given amount of stored energy in the quarter-wave mode, the voltage at the taps is $\sin (\pi/2)d_1/d$ times the voltage at the open end. Thus the impedance of each element in the equivalent circuit of the tap connection is $\sin^2 (\pi/2)d_1/d$ times the corresponding impedance in the equivalent circuit for connections made at the open end. The change in impedance level is taken into account in Fig. 28B by connecting the R-L-C circuit to the input terminals through an ideal transformer having a turns ratio of $\sin (\pi/2)d_1/d : 1$.

If d_1 is extremely small relative to d, the peak of the $|Z|$ curve corresponding to quarter-wave resonance of the entire line occurs at a frequency very close to the frequency of the dip resulting from quarter-wave resonance of the line section of length $d - d_1$, and coupling reactance is important. Because the dip occurs at a higher frequency than the peak, the coupling reactance is inductive. The coupling inductance, as indicated in Fig. 28B, is the difference between ld_1, the low-frequency induct-

ance measured at the taps, and $L \sin^2 (\pi/2)d_1/d$, the inductance of the equivalent circuit at resonance. For small values of d_1, $L \sin^2 (\pi/2)d_1/d$ is small compared with ld, and the low-frequency inductance of the line ld_1 may be taken as the coupling inductance with reasonable accuracy.

Connection to a resonant line may also be made inductively by means of a transformer, as shown in Fig. 29*A*. A peak-dip pair results from this method of connection in the following manner: With input terminals open, resonance occurs at a frequency f_1, for which the line length d is slightly less than a quarter wave. The line and its terminating inductance L_2 are then equivalent to a quarter-wave line open at the far end. With the input terminals short-circuited, resonance occurs in the same way except that the line is terminated in the short-circuit inductance of

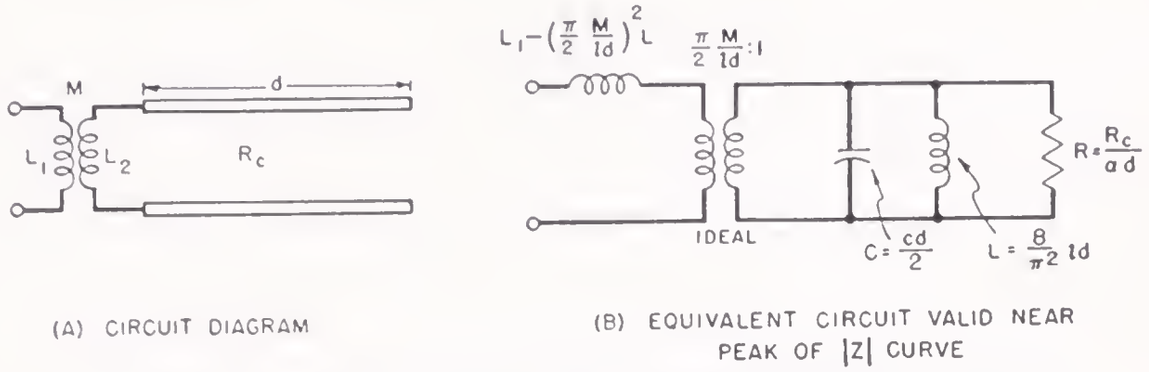


FIG. 29. Transmission-line resonator with transformer coupling.

the transformer $(1 - k^2)L_2$. Because $(1 - k^2)L_2$ is smaller than L_2, the short-circuit mode occurs at a slightly higher frequency f_2 for which d is more nearly a quarter wavelength. If L_2 is small, f_1 and f_2 are nearly equal and the input impedance curve of the system has a closely spaced peak-dip pair, the peak occurring at f_1 and the dip at f_2.

An equivalent circuit emphasizing the peak is simply obtained if L_2 is considered to be essentially a short circuit and d very nearly a quarter wavelength. The peak current in L_2 may then be taken as $I_{m_{\max}}$, and the peak voltage at the transformer input terminals is then

$$E_m = \omega_0 M I_{m_{\max}} = \frac{\omega_0 M}{R_c} E_{m_{\max}}$$

where M is the mutual inductance of the transformer. The fact that the voltage at the input terminals is $\omega_0 M/R_c$ times the voltage at the open end of the line indicates an equivalent circuit composed of a parallel R-L-C circuit based upon open-end quantities and an ideal transformer of turns ratio $\omega_0 M/R_c$. Coupling reactance must also be included, however, because of the presence of the peak-dip pair. The equivalent circuit thus takes the form of Fig. 29*B*, and R, L, and C have the values indicated in Fig. 26*D* for $n = 1$.

Setting ω_0 equal to $2\pi f_0$ and making use of the value of $f_0 = f_1$ in Eq. 32, p. 650, yields for the ideal-transformer ratio

$$\frac{\omega_0 M}{R_c} = 2\pi \frac{M}{4d\sqrt{lc}}\sqrt{\frac{c}{l}} = \frac{\pi}{2}\frac{M}{ld}$$

or $\pi/2$ multiplied by the ratio of M to the low-frequency inductance of the line. An approximate value for the coupling inductance is determined by subtracting from the low-frequency inductance L_1 the resonant-circuit inductance L referred to the transformer primary. The result, $L_1 - (\pi M/2ld)^2 L$, is indicated in the figure.

Comparison of Figs. 28*B* and 29*B* shows that tap connection coupling and transformer coupling used with transmission-line resonators provide very similar results. Either coupling method furnishes a means of controlling the resonator impedance. The impedance is decreased by moving the taps in Fig. 28 toward the shorting bar or by reducing the mutual inductance of the transformer in Fig. 29. With either coupling device, an excessive reduction of impedance may make the resonator impedance so small relative to the coupling reactance that the effect of resonance is lost.

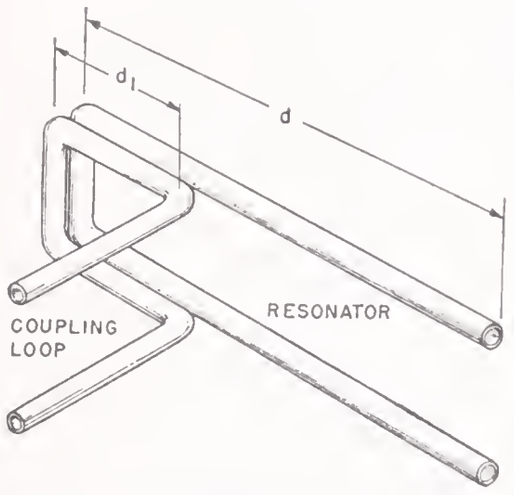


FIG. 30. Transmission-line resonator with coupling loop.

The similarity between the tap connection and the transformer may be further illustrated by Fig. 30. This figure also serves to point out that the windings of a transformer employed with high-frequency lines do not have many turns. Ordinarily the transformer comprises a single-turn coupling loop linked by mutual magnetic flux to the shorting bar of the resonator line. The particular coupling-loop arrangement shown in Fig. 30 is negligibly different from the tap connection, provided the loop is placed very close to the resonator line so that essentially the same flux links the loop and the d_1 section of the line. With this arrangement the loop voltage equals the voltage that would appear between taps on the resonator conductors at the distance d_1 from the short circuit.

Figure 29*B* applies to the resonator of Fig. 30. In this equivalent circuit, the ideal-transformer ratio $\pi M/2ld$ may be decreased (1) by reducing the length d_1 of the coupling loop in Fig. 30, the loop being kept close to the resonator, or (2) by moving the loop away from the resonator and not changing d_1. If the first method is employed, the loop inductance L_1 is decreased. Thus both the coupling reactance and the resonator imped-

ance decrease, but the resonator impedance decreases at the greater rate. If the second method is used, L_1 is not changed. The resonator impedance decreases, but the coupling reactance (see Fig. 29*B*) increases somewhat and then (for small values of M) remains constant. For a given resonant impedance, coupling reactance is least when the input line is coupled as closely as possible to the resonator or, in effect, when the tap connection is used.

A resonant system that illustrates one way of employing capacitive coupling is shown in Fig. 31*A*. Connections are made to metal disks which serve as small capacitors connecting the external circuit to the

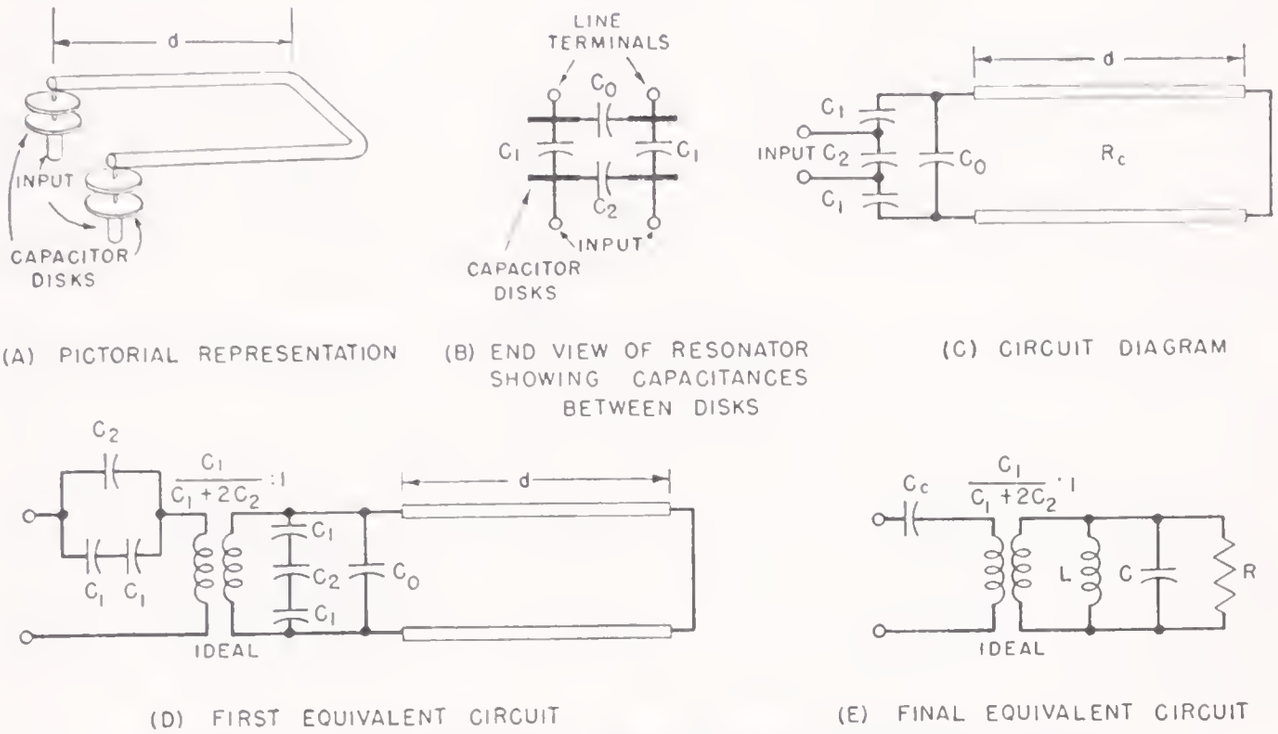


FIG. 31. Transmission-line resonator with capacitive coupling.

resonant line. The important capacitances between disks are indicated in Fig. 31*B*, and a circuit diagram showing the line and the capacitances is shown in Fig. 31*C*. The lowest frequency open-circuit mode of the system occurs at a frequency for which d is slightly less than a quarter wave and corresponds to the lowest frequency mode of the capacitor-loaded line of Fig. 27, p. 653.

An equivalent circuit applicable in the frequency range of the lowest frequency open-circuit mode may be obtained if the capacitance voltage divider of Fig. 31*C* is first replaced by an ideal transformer with shunt and series capacitors (see Art. 20, Chap. II). The result is Fig. 31*D*, and the portion of this circuit to the right of the ideal transformer is changed to the *R-L-C* circuit of Fig. 31*E* by the methods employed in connection with Fig. 27. Note that the capacitive coupling reactance of Fig. 31*E* indicates a dip in the $|Z|$ curve at a frequency less than the frequency of the peak. The dip is the result of the short-circuit mode in which the

line of Fig. 31*C* resonates with C_0 shunted by $\frac{1}{2}C_1$, and the peak is the result of the open-circuit mode obtained at the slightly higher frequency of resonance of the line with C_0 shunted by the series combination of $\frac{1}{2}C_1$ and C_2.

10. Transmission Lines Terminated by Resonators. In the uhf and microwave-frequency ranges, resonators are oftentimes connected to other apparatus through a short section of transmission line. Because of the properties of transmission lines, the impedance at the input terminals of the line may be very different from the input impedance of the resonator itself. The presence of the transmission line must therefore be taken into account in any equivalent-circuit representation of the entire system.

Figure 32*A* shows several different lengths of transmission line terminated by a parallel *R-L-C* circuit. The *R-L-C* circuit may be regarded as an equivalent circuit for any type of resonator in the region of an isolated peak of its impedance curve. The peak is assumed to be located at a frequency f_0, and the values of d in the figure represent line lengths at this frequency.

The variation of the input impedance of the resonator-and-line system as a function of frequency is indicated by the impedance curves of Fig. 32*B*. The first-row curve, drawn for $d = 0$, is the usual $|Z|$ curve for a parallel-resonant circuit and indicates that the impedance is essentially zero except at frequencies near f_0. The remaining impedance curves therefore follow the input-impedance curve for a shorted-end line, except in the range of frequencies near f_0. At f_0 the impedance is that of the line terminated in a high resistance.

The impedance curve in the third row of Fig. 32, applicable for $d = \lambda/4$ at f_0, is the curve for a shorted-end line near its quarter-wave resonance, except that a sharp dip occurs at the frequency f_0. The high terminating impedance at this frequency causes the quarter-wave line to have a low input impedance. Similarly, for $d = \lambda/2$, the curve in the bottom row of the figure indicates a peak of impedance at f_0.

For a line length $d = \lambda/8$, the impedance curve has a peak-dip pair in the neighborhood of f_0, as indicated by the second curve in Fig. 32*B*. The peak occurs at f_1, a frequency slightly less than f_0, and corresponds to an open-circuit mode of the *R-L-C* circuit shunted by the capacitance of the $\lambda/8$ section of open-end line. The dip at a frequency f_2 above f_0 is the result of a short-circuit natural mode obtained when the *R-L-C* circuit is shunted by the inductance of the $\lambda/8$ section of shorted-end line. For $d = 3\lambda/8$, the curve in the fourth row applies and shows a similar peak-dip pair. The dip in this figure, however, is at a lower frequency than the peak because the $3\lambda/8$ line is equivalent to a capacitance when short-circuited and to an inductance when open.

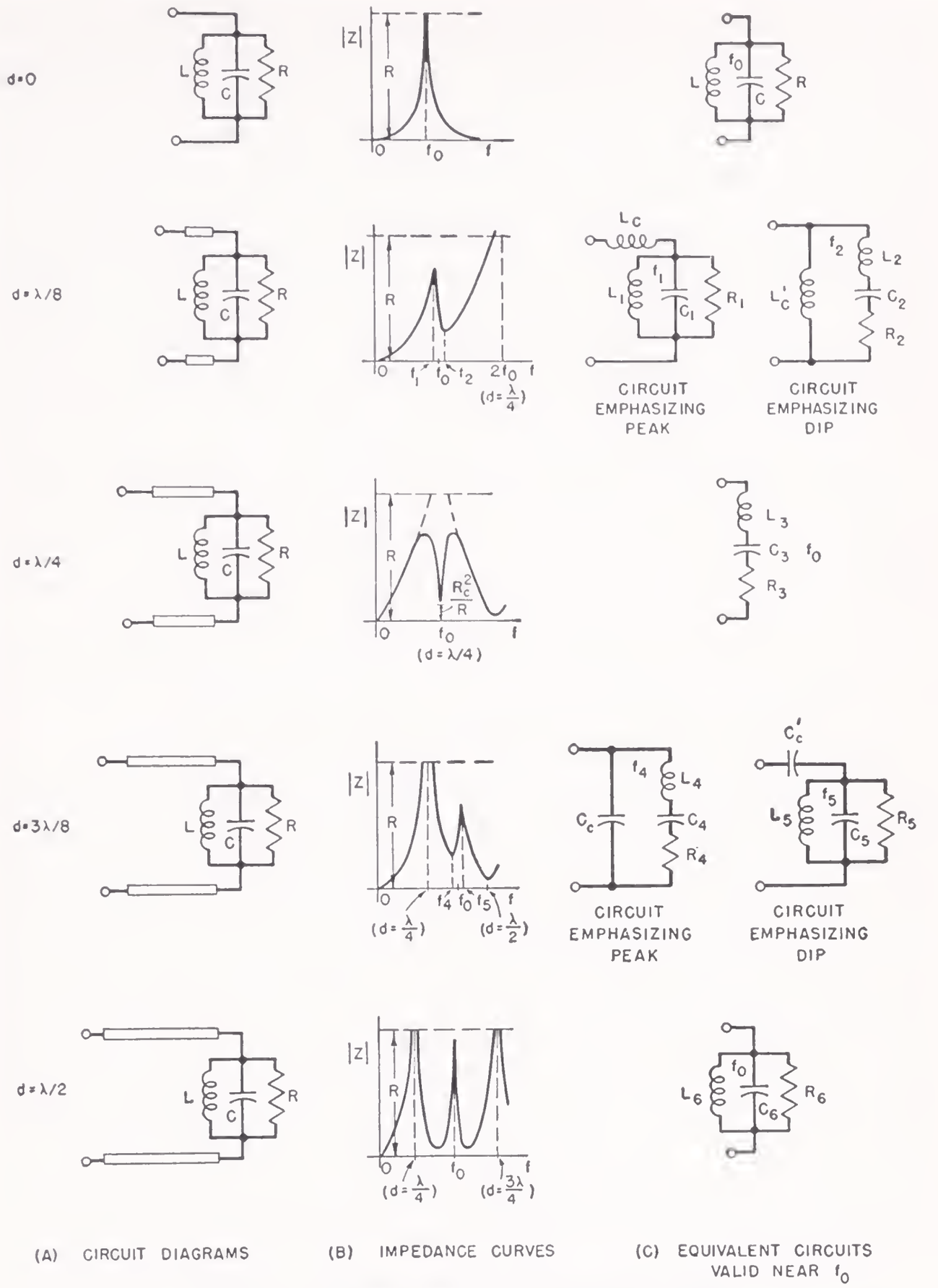


FIG. 32. Resonator with input transmission line.

From these impedance curves appropriate forms for equivalent circuits valid near f_0 can be determined to represent the resonator-and-line system of Fig. 32. If the input line is an even number of quarter waves long at f_0, a peak in the impedance curve must be represented. A parallel R-L-C combination is thus required in the equivalent circuit, as indicated

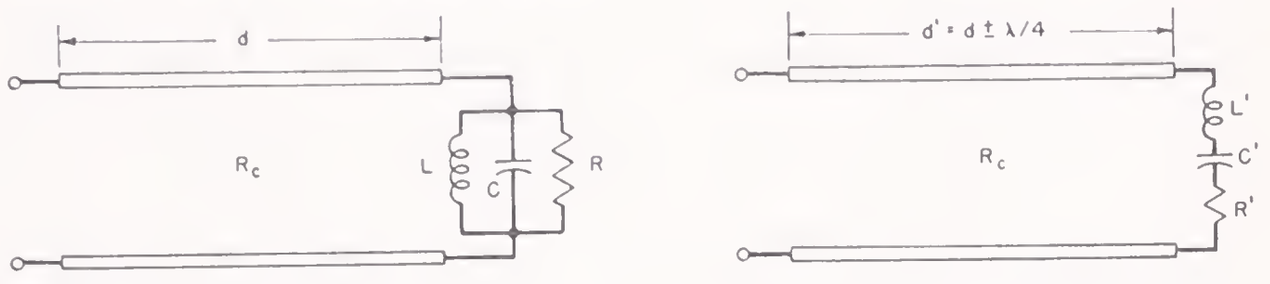
by the top and bottom diagrams of Fig. 32*C*. If the line is an odd number of quarter waves in length at f_0, the dip in the impedance curve requires a series *R-L-C* combination in the equivalent circuit (see the center diagram of Fig. 32*C*). For intermediate line lengths, peak-dip pairs occur in the impedance curve and the equivalent circuit comprises a parallel tuned circuit in series with a coupling reactance or a series tuned circuit shunted by a coupling susceptance. As indicated in Fig. 32*C*, the coupling reactance (or susceptance) is inductive if the peak occurs at a lower frequency than the dip and capacitive if the dip is at the lower frequency.

The parameters of each of the resonant circuits in Fig. 32*C* may be calculated from the natural frequency, stored energy, and Q of the resonant mode represented. The values obtained differ from the values of R, L, and C of the terminating resonator itself. This change of element values is to be expected in all the circuits of Fig. 32*C* except, possibly, in the one for $d = \lambda/2$ because of the changes in resonant frequency and form of the circuit that occur. Even for $d = \lambda/2$, changes of element values occur, however, because C_6 (see Fig. 32*C*) must be greater than C to account for stored energy in the input line and R_6 must be less than R to account for losses in the line. The Q of the equivalent circuit may be greater or less than that of the terminating resonator, depending upon the relation of stored energy to loss in the input line.

Ordinarily the impedance-level parameter $\sqrt{L/C}$ of the terminating resonator is small relative to the characteristic resistance of the line, and the line length does not exceed a few wavelengths. The peaks and dips of the impedance curve caused by the resonator-impedance changes are then very sharp relative to the peaks and dips produced by the line itself, and the stored energy and loss in the line are negligible relative to the stored energy and loss in the resonator. In consequence, for a half-wave input line the equivalent-circuit parameters (R_6, L_6, and C_6 in Fig. 32*C*) are essentially equal to the corresponding terminating-resonator parameters. For a quarter-wave input line of characteristic resistance R_c, the input impedance is very nearly R_c^2 divided by the resonator impedance. Thus in Fig. 32*C*, R_3 is almost equal to R_c^2/R and $\sqrt{L_3/C_3}$ is approximately $R_c^2/\sqrt{L/C}$. For values of d that are not multiples of $\lambda/4$, the parallel-resonant part of an equivalent circuit emphasizing the peak of the $|Z|$ curve differs from the terminating resonator mainly in that the impedance level is higher. For resonance in the open-circuit mode, the standing-wave pattern of the input line has the effect of an ideal transformer and causes the voltage at the terminating resonator to be less than the input voltage of the line.

The equivalences between circuits indicated in Fig. 32 have been presented as a means of obtaining a lumped-parameter circuit equivalent to a transmission-line-and-resonator circuit. In the study of high-fre-

quency systems, however, these same equivalences are particularly useful when applied in the reverse direction. Because any one-entry resonator for which the impedance curve has an isolated peak, dip or peak-dip pair can be represented in a small frequency range near the resonance by one of the circuits of Fig. 32*C*, and because each of these circuits is equivalent to a transmission-line section terminated by a parallel *R*-*L*-*C* circuit, the line-and-resonator circuit (drawn in Fig. 33*A*) may be regarded as a general equivalent circuit for a one-entry resonator. As indicated by the third-row diagrams of Fig. 32, the resonator and last-quarter-wave section of line in Fig. 33*A* may be replaced by a series *R*-*L*-*C* circuit. Thus the circuit of Fig. 33*B* is an alternate to that of Fig. 33*A* as an equivalent circuit for the one-entry resonator.



(A) PARALLEL-CIRCUIT REPRESENTATION (B) SERIES-CIRCUIT REPRESENTATION

FIG. 33. Equivalent circuits for a one-entry resonator.

The equivalent circuits of Fig. 33 are especially useful at high frequencies, because connections to high-frequency resonators are almost always made through transmission lines, and the exact location of the resonator input terminals is difficult to determine when wavelengths are comparable to the line and resonator dimensions. When one of the equivalent circuits of Fig. 33 is employed, it is necessary only to select an arbitrary reference plane in the input line. The remainder of the line and the resonator are then represented by either of the circuits in Fig. 33, parameter values being chosen in accord with the input impedance at the chosen reference plane. A change in the location of the reference plane is accounted for by a corresponding change of the line length in the equivalent circuit. In fact, if desired, a reference plane may be chosen for which the line length in one of the equivalent circuits is zero. The input impedance at the reference plane is then the impedance of a simple parallel or series *R*-*L*-*C* circuit.

In order to specify either of the equivalent circuits of Fig. 33, four parameters—*d*, *R*, *L*, and *C* for the circuit of Fig. 33*A*—are required. These quantities are most conveniently determined from a different set of four parameters that is more directly related to the measurements that can be made upon the system. The second set of parameters comprises the *input-line length, resonant frequency, unloaded Q*, and *loaded Q*. The

input-line length is the distance d or d' in Fig. 33. The resonant frequency (f_0) and unloaded Q (denoted by Q) apply to the terminating resonator alone and are the parameters introduced in Arts. 1 and 2. The loaded Q (denoted by Q_l) is the Q of the circuit formed when the characteristic resistance of the transmission line is connected to the resonator terminals. The loaded Q may be thought of as the Q obtained when oscillations of the resonators in Fig. 33 are damped not only by the resonator losses but also by loss of power to the transmission line, the line length being assumed infinite. For the parallel-resonator circuit in Fig. 33*A*,

$$f_0 = \frac{1}{2\pi\sqrt{LC}}, \qquad Q = \frac{R}{\sqrt{L/C}}, \qquad Q_l = \frac{RR_c/(R + R_c)}{\sqrt{L/C}} \tag{40}$$

and for the series-resonator circuit of Fig. 33*B*,

$$f_0 = \frac{1}{2\pi\sqrt{L'C'}}, \qquad Q' = \frac{\sqrt{L'/C'}}{R'}, \qquad Q_l' = \frac{\sqrt{L'/C'}}{R' + R_c} \tag{41}$$

The parameter Q_l ordinarily replaces the impedance-level parameter $\sqrt{L/C}$ introduced in Art. 6 whenever a resonator is used in conjunction with a transmission line, because Q_l specifies the relationship of $\sqrt{L/C}$ to the characteristic resistance of the line and determines the sharpness of the resonance effect in the over-all system. For the parallel-resonator circuit of Fig. 33*A*, low values of $\sqrt{L/C}$ relative to R_c lead to values of Q_l only slightly less than Q and result in a terminating impedance for the transmission line that changes between nearly zero and nearly infinite values in a narrow frequency range near f_0. High values of $\sqrt{L/C}$ make Q_l low and cause the line to be terminated almost in an open circuit at all frequencies near f_0. For the series-resonator circuit of Fig. 33*B*, the impedance relationships are reversed; large values of $\sqrt{L'/C'}$ relative to R_c make Q_l large and provide a sharply resonant over-all system. Often a resonator is said to be *loosely coupled* to a line when the impedance relations are such that Q_l is nearly as large as Q and *closely coupled* if Q_l is small and Q large. (This use of "loosely coupled" and "closely coupled" should not be confused with the concept of degree of coupling in a transformer, which is expressed by the coupling coefficient.)

The parameters f_0, Q, Q_l, and d for a one-entry resonator can be determined from standing-wave measurements in the input line. Again consider the line and parallel *R-L-C* circuit in Fig. 33*A*. The reflection coefficient Γ_L at the end of the line is given by the expression

$$\Gamma_L = \frac{1 - y_L}{1 + y_L} = \frac{1 - g - jb}{1 + g + jb}$$

where $y_L = g + jb$ is the per-unit admittance of the terminating resonator, and thus g is equal to R_c/R and b equals $R_c(\omega C - 1/\omega L)$. Figure 34 is a plot of Γ_L on the Smith chart for frequencies both above and below f_0. At f_0, Γ_L is located at the point A, which corresponds to

$$\frac{1 - g}{1 + g} = \frac{R/R_c - 1}{R/R_c + 1}$$

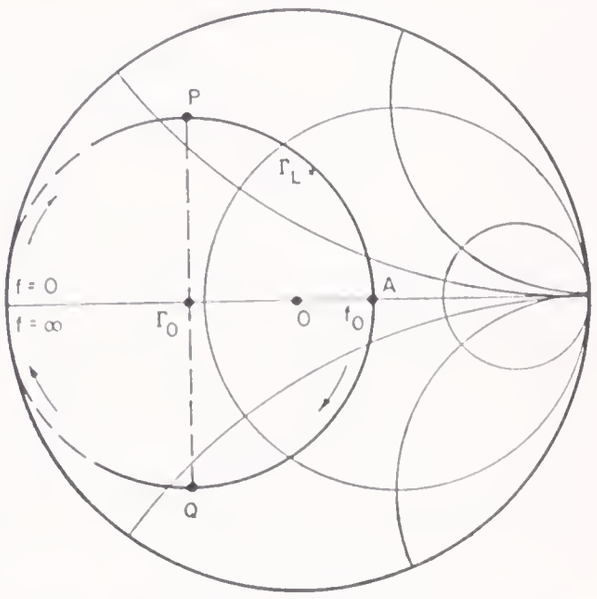


FIG. 34. Locus described by Γ_L as frequency is varied.

As the frequency varies, Γ_L varies in a manner which causes it to trace the circular path on the chart.

That Γ_L describes a circle as the frequency is varied may be proved by showing that a point exists which is equally distant from all points of the locus of Γ_L. The location of such a point Γ_0, which is the center of the circle, may be established in the following way: The distance between Γ_0 and a point Γ_L on the locus is the magnitude of the complex number

$$\Gamma_L - \Gamma_0 = \frac{1 - g - jb}{1 + g + jb} - \Gamma_0 = \frac{(1 - g) - \Gamma_0(1 + g) - jb(1 + \Gamma_0)}{(1 + g) + jb}$$

or

$$\Gamma_L - \Gamma_0 = (1 + \Gamma_0)\frac{(1 - g) - \Gamma_0(1 + g) - jb(1 + \Gamma_0)}{(1 + g)(1 + \Gamma_0) + jb(1 + \Gamma_0)} \tag{42}$$

Numerator and denominator of the fractional factor in Eq. 42 have equal-magnitude imaginary parts. If the real parts are made equal, the fraction will have a magnitude of unity and $|\Gamma_L - \Gamma_0|$ will equal $|1 + \Gamma_0|$, a constant. Thus the locus is proved to be a circle with Γ_0 as its center if Γ_0 can be located so that

$$(1 - g) - \Gamma_0(1 + g) = (1 + \Gamma_0)(1 + g)$$

The value of Γ_0 that satisfies this equation is $-g/(1 + g)$, and the center of the circular locus is located at the point marked Γ_0 in Fig. 34. The radius of the circle, $1 + \Gamma_0$, is equal to $1/(1 + g)$. Substituting the value of Γ_0 into Eq. 42 yields

$$\Gamma_L - \Gamma_0 = \frac{1}{1 + g}\frac{(1 + g) - jb}{(1 + g) + jb} \tag{43}$$

It may be shown also that points P and Q in Fig. 34 correspond to the half-power frequencies of the circuit from which Q_l is computed, that is,

the parallel R-L-C circuit of Fig. 33A shunted by R_c. At points P and Q, the angle of $\Gamma_L - \Gamma_0$ is, respectively, 90 and -90 deg. According to Eq. 43, the angle of $\Gamma_L - \Gamma_0$ is ± 90 deg when $1 + g = \mp b$, that is, when

$$1 + \frac{R_c}{R} = \mp R_c \left(\omega C - \frac{1}{\omega L}\right)$$

or

$$\frac{1}{R_c} + \frac{1}{R} = \mp \left(\omega C + \frac{1}{\omega L}\right) \tag{44}$$

As explained in Art. 6 of Chap. VI, Eq. 43 represents the relation between conductances and susceptances at the half-power frequencies of a circuit composed of L, C, R, and R_c in parallel. Therefore the per-unit bandwidth between half-power frequencies of this circuit is equal to $1/Q_l$, the reciprocal of the loaded Q of the resonator.

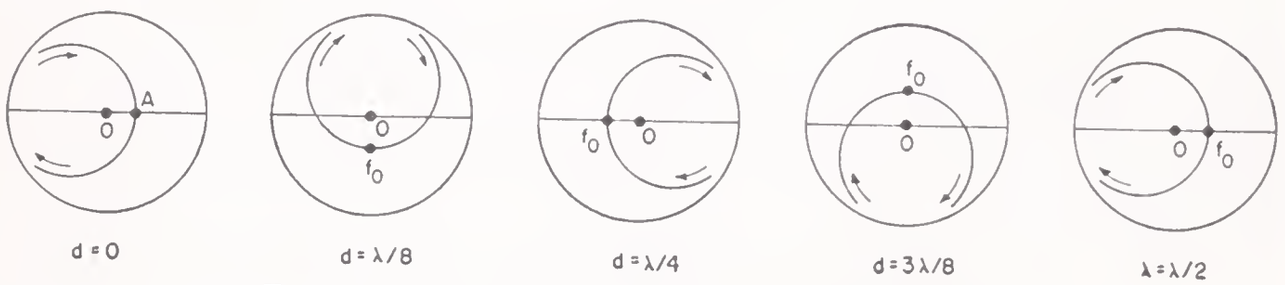


FIG. 35. Locus of Γ_d for the line lengths of Fig. 32.

The locus of the reflection coefficient Γ_d at a reference plane distant d from the R-L-C circuit may be determined from the Γ_L circle by rotating each point of the locus clockwise through an angle equal to twice the electrical length of the line between load and reference plane. Because each point of the Γ_L circle corresponds to a different frequency, each point should be rotated through a slightly different angle. If Q_l is high, however, most of the circle corresponds very nearly to the frequency f_0, and a close approximation to the Γ_d locus is obtained by rotating the entire circle through a constant angle equal to twice the electrical length of the line at f_0. The resulting error is appreciable only at frequencies outside the range of interest.

Figure 35 shows the positions of Γ_d circles for the line lengths considered in Fig. 32. Observe that, if d is a multiple of a half wave, the center of the circle lies to the left of the center of the chart, and frequency f_0 (corresponding to point A) represents a peak in the input-impedance curve. If d is an odd multiple of a quarter wavelength, the center of the Γ_d locus lies to the right of the center of the chart, and f_0 corresponds to a dip in the impedance curve. A circle displaced above or below the center of the chart corresponds to a line length which is not a multiple of a quarter wave and to a peak-dip pair in the impedance curve. Frequency f_0 then

corresponds to the oscillation frequency of the resonator itself, the peak-and-dip frequencies being slightly different from f_0.

Curves similar to those of Fig. 35 may be plotted from standing-wave measurements made in the input line of any one-entry resonator, and from these curves the parameters R, L, C, and d of the equivalent circuit may be determined. Observe in Fig. 35 that the lowest standing-wave ratio always occurs at f_0. The resonant frequency of the R-L-C circuit can therefore be determined directly. The value of R required to produce the minimum standing-wave ratio may be calculated from knowledge of the characteristic resistance of the line. The loaded Q is determined from frequency difference between points corresponding to P and Q in Fig. 34, and from Q_l, f_0, and R the values of L and C can be calculated. The

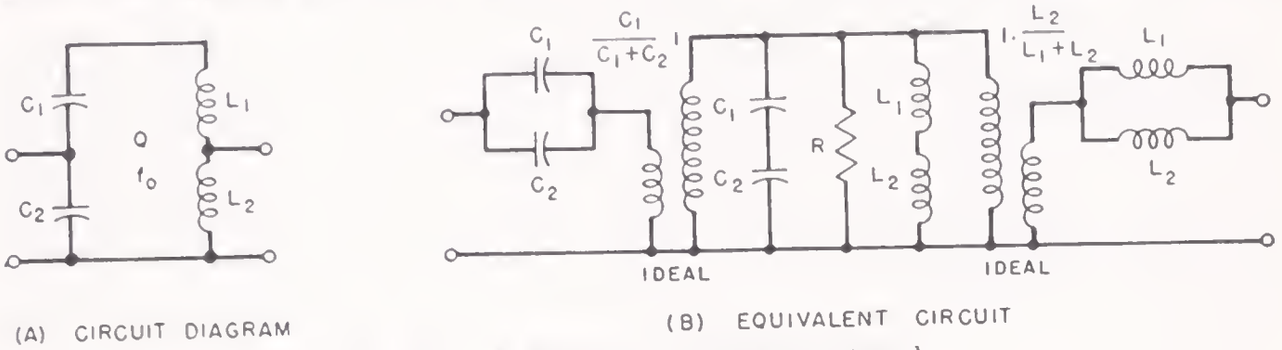


FIG. 36. A two-entry resonant network.

value of d is determined from the position of the Γ_d circle in relation to the circle for $d = 0$.

11. Two-entry Resonators. Frequently power delivered to a resonator at one point of entry is extracted from the resonator at another point of entry. Each entry may be a pair of input terminals or, in the microwave region, a transmission line or waveguide. Such a resonator is called a two-entry resonator.

A simple form of two-entry resonator, consisting of a parallel L-C network and two pairs of terminals, is drawn in Fig. 36A. A source is connected across one of the two capacitors, and the load is placed across one of the two inductors.

An equivalent circuit is drawn in Fig. 36B. Resistance R accounts for any small loss that may be present in the original circuit, and the ideal transformers and coupling reactances represent the effect of the capacitance and inductance voltage dividers (see Art. 20, Chap. II).

The operation of the circuit at resonance may be visualized readily if the current in R and the voltages across the coupling reactances are temporarily ignored. Since the inductance and capacitance branches together constitute an open circuit at resonance, the resonator is simply the equivalent of an ideal transformer of ratio

$$\frac{\text{Output voltage}}{\text{Input voltage}} = \frac{L_2/(L_1 + L_2)}{C_1/(C_1 + C_2)}$$

connecting input and output terminals. For this reason such a two-entry resonator is sometimes called a *resonant transformer*. The taps on the voltage dividers control the transformation ratio and thus serve to adjust impedance relationships between the input and output circuits.

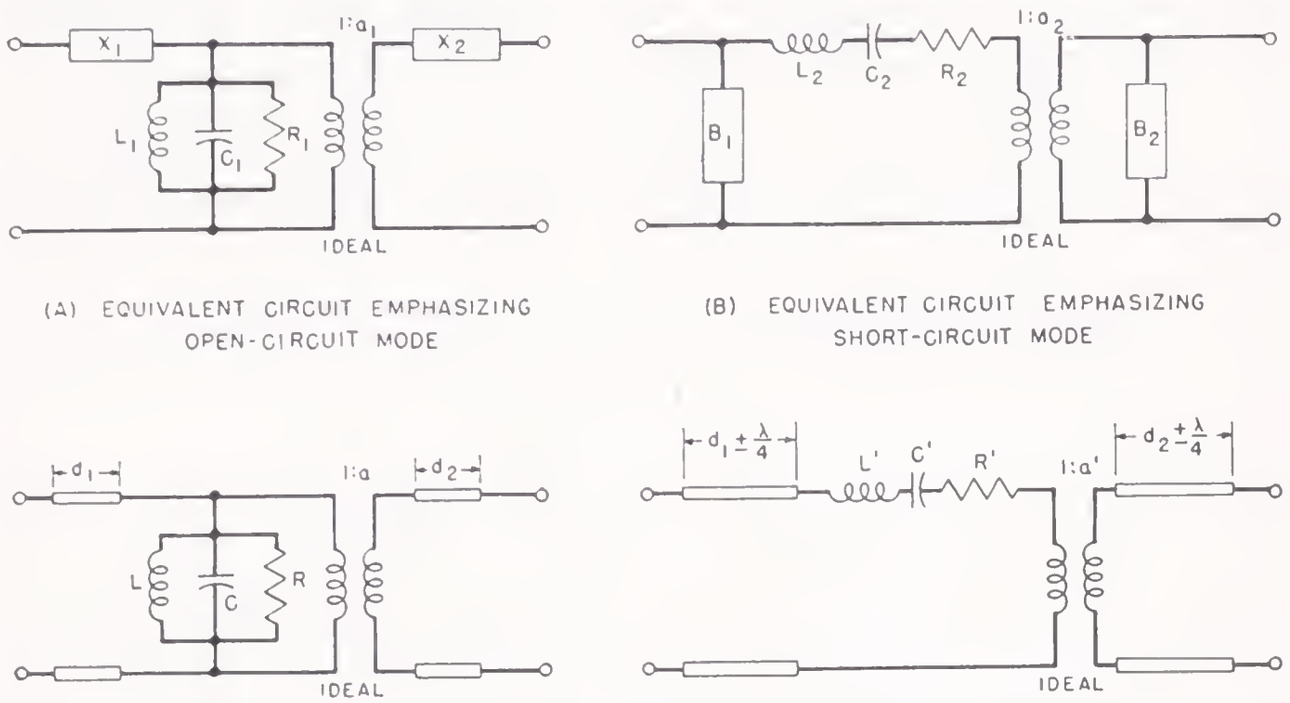
Actually, of course, the operation of the resonator is influenced by the shunt resistance R, especially when the voltage-divider taps are adjusted for high voltages across the resonant circuit so that the loss in R is important. Actually, also, if appreciable load current is drawn, the voltage drops across the coupling reactances may become important. From the point of view of operation at resonance, therefore, the taps should be adjusted to make both $C_1/(C_1 + C_2)$ and $L_2/(L_1 + L_2)$ as near unity as possible, not only to minimize the loss in the resonator, but also to reduce the magnitudes of the coupling reactances (see Art. 7).

At frequencies above and below resonance, the reactance of the tuned circuit is extremely low compared with its value at resonance, and the source and load are essentially short-circuited, except for coupling reactances. The resonator may therefore be likened to a frequency-operated switch which connects source to load only in a small band of frequencies centered at f_0. If an extremely narrow pass band is desired, the voltage-divider taps should be adjusted for very small values of $C_1/(C_1 + C_2)$ and $L_2/(L_1 + L_2)$ in order that the resonator may short-circuit the input and output terminals effectively at frequencies only slightly different from f_0. The adjustment chosen for the taps is therefore a compromise between sharp selectivity and high efficiency at resonance.

The resonator of Fig. 36*A* is characterized by a single natural mode of resonance with both terminals open and by the fact that the mode is coupled to *both* terminal pairs. Although resonators in general may have a variety of resonant modes, usually only those modes within a narrow frequency band which couple to both pairs of terminals are of interest. An equivalent circuit that applies to any two-entry resonator in the region of a single open-circuit mode is drawn in Fig. 37*A*. The tuned-circuit elements R_1, L_1, and C_1 are determined by the resonant frequency, Q, and stored energy of the open-circuit mode in accordance with the methods of Art. 6, and the ideal transformer with ratio $1:a_1$ accounts for the different voltage and impedance levels that are present at input and output terminals. If desired, the transformer may be placed on the opposite side of the tuned circuit from that shown or two transformers may be used as in Fig. 36*B*. The coupling reactances account for modes obtained at near-by frequencies with one or both terminal pairs short-circuited. Approximate values of these reactances may be determined from the inductance or capacitance measured between terminals at very low frequencies (see Art. 9).

The equivalent circuit of Fig. 37*A* assumes that the entire coupling between input and output terminals is provided by the open-circuit mode of interest. If other coupling is present, a mutual coupling reactance must be placed in series with the tuned circuit in order to provide a connection between source and load at frequencies for which the tuned circuit is essentially a short circuit. In practical resonators, mutual coupling reactance is negligible for most purposes, and the circuit of Fig. 37*A* is adequate.

The equivalent circuit of a two-entry resonator in the region of an isolated short-circuit mode that couples to both entries is drawn in Fig.



(A) EQUIVALENT CIRCUIT EMPHASIZING OPEN-CIRCUIT MODE

(B) EQUIVALENT CIRCUIT EMPHASIZING SHORT-CIRCUIT MODE

(C) EQUIVALENT CIRCUITS OF PARTS A AND B WITH COUPLING REACTANCES REPLACED BY TRANSMISSION LINES

FIG. 37. Equivalent circuits for a two-entry resonator.

37*B*. This circuit is similar to the series-resonant equivalent circuit of a one-entry resonator, except that it includes two terminal pairs and an ideal transformer to account for any change in voltage and impedance level in the resonator. The series *R-L-C* circuit is resonant at the natural frequency determined with both terminal pairs short-circuited, and the coupling reactances account for near-by natural frequencies obtained with one or both terminal pairs open. If nonresonant coupling exists, it may be represented by a shunt susceptance placed across the *R-L-C* combination.

At high frequencies, input and output connections to resonators are usually made by means of transmission lines. Therefore, as for one-entry resonators, it is convenient to use equivalent circuits in which input and output line sections replace the coupling reactances and susceptances of Figs. 37*A* and *B*. Appropriate changes of the values of resonant-circuit parameters and the transformer ratio are required, of course, and the

resulting circuits are indicated in Fig. 37*C*. *Either* of the two circuits shown may be used to represent *any* two-entry resonator of the general type considered. The equivalence of the two circuits may be shown from the fact that a parallel *R-L-C* circuit joined by two transmission lines to two terminal pairs has a short-circuit natural frequency at f_0, the resonant frequency of the *R-L-C* circuit, provided each line is a quarter wave in length at f_0. Therefore the parallel circuit with quarter-wave lines is equivalent to a two-entry circuit in which a series *R-L-C* circuit resonant at f_0 joins the terminal pairs.

For a two-entry resonator, loaded Q has the same basic significance as for the one-entry resonator; that is, Q_l is the Q of the circuit obtained from the equivalent circuit by replacing each line with its characteristic resistance. For example, if each line of the circuit of Fig. 37*C* employing the parallel resonator has the characteristic resistance R_c, Q_l is the Q of the circuit comprising L, C, R, R_c, and R_c/a^2 all in parallel. The loaded Q indicates the sharpness of resonance of the resonator when operated in conjunction with the lines.

12. Coupling Devices for High-frequency Resonators. The equivalent circuits discussed in Arts. 6–11 apply directly to resonators and coupling

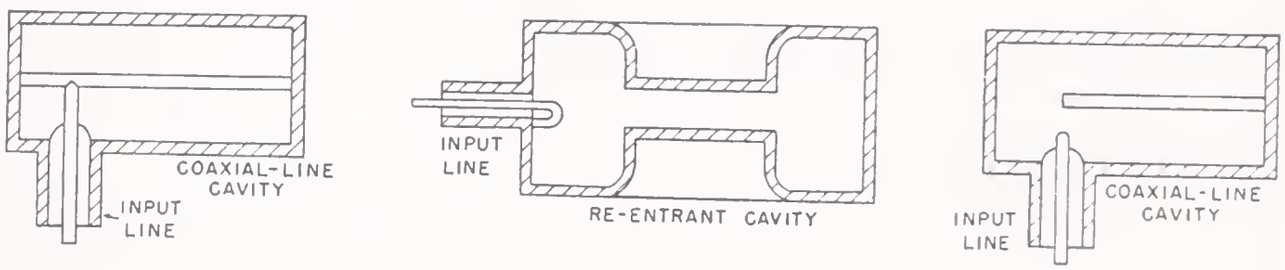


FIG. 38. Methods of connecting a coaxial line to a cavity resonator.

devices operating at any frequency in the range from very low to ultrahigh frequencies. In low-frequency network resonators, definite terminal pairs can be selected at which voltage, current, and impedance can be measured, and at ultrahigh frequencies definite reference planes in input or output lines can be selected at which the same quantities can be measured. At higher frequencies, waveguides and cavity resonators are employed, and the operation of these devices is described in terms of field intensities and wave impedances. The relation of these quantities to the circuit quantities and equivalent circuits for high-frequency resonators are to be the subjects of the remaining articles in this chapter.

Microwave resonators may be joined to other apparatus by means of either coaxial lines or waveguides. A line or guide may be coupled to a resonator in a variety of ways. For coaxial lines, three methods of coupling to a cavity resonator are indicated in Fig. 38. If the resonator is a coaxial-line cavity, *junction coupling* obtained by a direct connection of

the coaxial conductors may be employed, as indicated in Fig. 38*A*. This coupling method is the high-frequency analogue of the connection to taps on transmission-line conductors considered in Art. 9. *Loop coupling*, illustrated in Fig. 38*B*, is obtained if the inner conductor of the input line is bent around and fastened to the cavity wall. *Probe coupling*, illustrated in Fig. 38*C*, results if the inner conductor merely penetrates a short distance into the cavity. Loop and probe coupling are, respectively, the high-frequency analogues of the inductive and capacitive coupling methods discussed in Art. 9.

One method of coupling a waveguide to a cavity resonator makes use of a short section of coaxial line between guide and cavity. One end of the line connects to the cavity by one of the methods of Fig. 38, and the other end is coupled to the guide in somewhat similar fashion. Details of waveguide-to-coaxial-line junctions are described in Art. 17. A mechanically simpler coupling method for a waveguide and cavity resonator, illustrated in Fig. 39*A*, employs an *aperture*, or window, cut in a wall that is common to the guide and resonator. Coupling results from the penetration of electric and magnetic fields through the opening. As

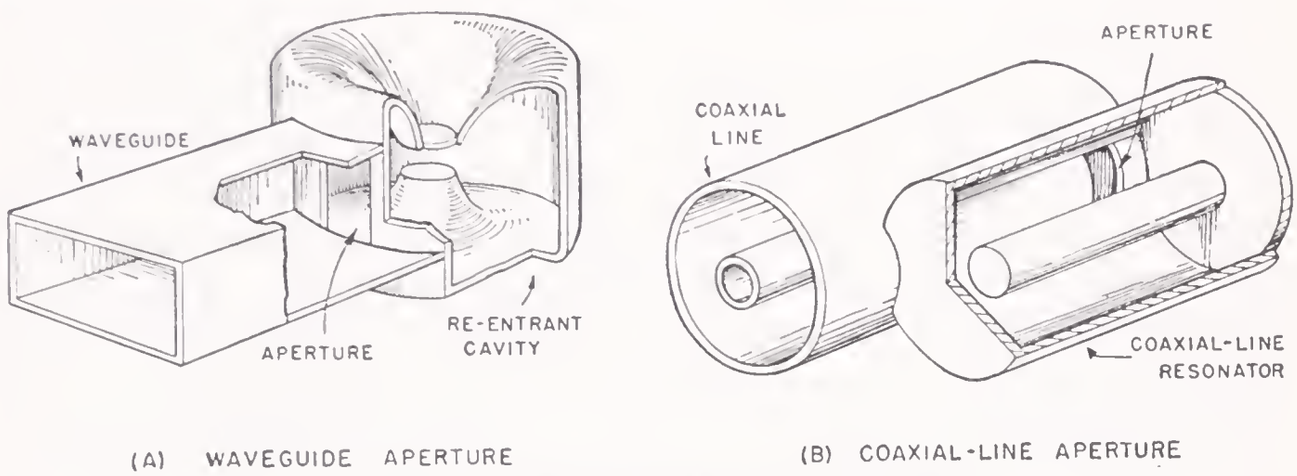


FIG. 39. Examples of aperture coupling.

indicated in Fig. 39*B*, aperture coupling may be employed in coaxial-line systems also.

No matter which of these coupling methods is employed, a cavity resonator with its input line or guide can be represented in the frequency region of an isolated mode by an *R-L-C* circuit connected to a transmission line. Any coupling reactance that may be present is taken into account by the length of the line in the equivalent circuit. By means of standing-wave measurements in the input line or guide, the parameters of the equivalent circuit (referred to a particular plane) may be determined. The performance of the resonator in association with other equipment can then be calculated. As an aid to understanding the physical principles of the resonators and coupling devices, however, and as a means of determining the effect of specific changes in the system, more detailed equivalent

circuits are desirable. Such circuits are developed in the following articles.

13. Parallel Connections at High Frequency. The coaxial-line junction of Fig. 38*A* is only one of many microwave coupling devices that are representable in equivalent circuits by means of elements connected in *parallel.* An essential feature of a parallel connection is that the same voltage is impressed across each of the elements in parallel. Parallel connections are simply made at low frequency by connecting impedance elements across the same pair of terminals. At high frequency, however, such a connection does not necessarily subject the elements to the same voltage. The reason for this difference between low- and high-frequency operation lies in the basic concept of voltage and its application to high-frequency devices.

Voltage is properly defined as *the integral of the tangential component of electric-field intensity along a specified line.* In Fig. 40*A*, the dashed line connecting points *A* and *B* is parallel to a uniform electric field of intensity $\mathcal{E}$, and the voltage of *A* with respect to *B* is simply $\mathcal{E}$ times the length of the line *AB*. Where the field pattern is more complicated, as in Fig. 40*B*, the path along which the voltage is measured must be specified, and an integration is needed to find the voltage between two points. To determine the voltage of point *C* with respect to point *E* along the path *CDE*, for example, the path must be divided into small segments. The voltage across each segment is then found by multiplying the component of $\mathcal{E}$ tangential to the segment by the segment length, and all the segment voltages are summed. Evidently the voltage along paths *CDE* and some other path, *CFE* for example, are equal only if the integral of the tangential components of $\mathcal{E}$ around the closed loop formed by the two paths is zero, that is, only if the *electromotive force* acting around the closed loop is zero.

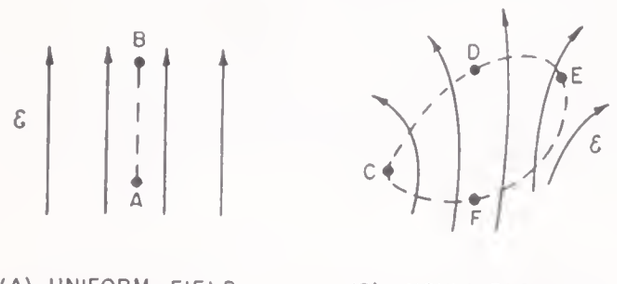


FIG. 40. Voltage between two points in a uniform and a nonuniform field.

As explained in Art. 17, Chap. VII, the electromotive force around any closed path is equal to the time rate of change of magnetic flux linking the loop. At low frequencies, the electromotive force is nearly always negligible (unless the loop is extremely large or encircles a ferromagnetic material with large flux density), and the voltage between any two points in the field may be specified without regard for the path between the points. At high frequencies, however, even a small magnetic flux linking the loop may induce appreciable electromotive force because of the high rate of change of flux. The path along which a voltage is measured thus

becomes important, and different voltages may be expected along different paths between the same two points.

Even though two impedances are connected to the same pair of points, they are in parallel only if the voltages along the two paths occupied by the impedances and their connecting leads are equal. At high frequencies, the voltage to which an impedance is connected must therefore be determined from the electric field in the region occupied by the impedance. It is not sufficient to know the points to which the impedance is connected.

Connection of impedance elements in parallel with a transmission line propagating only transverse-electromagnetic waves is of special interest. Because the magnetic field of the waves lies entirely in the transverse plane, it cannot link any closed loop located in that plane. Therefore, if

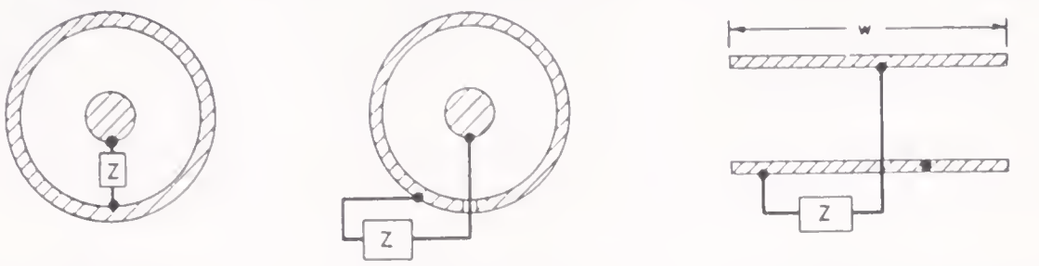


FIG. 41. Impedances in parallel with transmission lines.

an impedance element in the transverse plane is connected by leads in the same plane to the line conductors, the impedance is subjected to the line voltage at the point of connection and may be considered in parallel with the line. Of course, no physical impedance device is confined completely to a single plane, but as long as the dimensions of the device along the line are short relative to the wavelength, voltages measured along different paths through the device are nearly equal, and the parallel connection may be considered a good approximation to the correct representation of the device.

With parallel-wire or parallel-strip lines, an impedance element connected in parallel with the line ordinarily lies outside the region between conductors. No change in the electrical conditions results, however, if the impedance device and its connections lie along a straight line between the centers of the conductors. An equivalent location of an impedance in parallel with a coaxial line is indicated by the sketch in Fig. 41*A*. Such a location of an impedance element inside a coaxial line is ordinarily impractical, and the alternate arrangement of Fig. 41*B* is employed instead. One lead passes through a hole in the outer conductor and thus permits the impedance element to be located outside the line. An analogous connection for a parallel-strip line is indicated in Fig. 41*C*. Because these connections are merely particular arrangements of the cir-

cuit within the transverse plane, they are each represented in an equivalent circuit as a connection of an impedance in parallel with the line.

The impedance in parallel with the line includes not only the impedance of the device marked Z in Fig. 41 but also the effect of the series inductance and shunt capacitance of the connecting leads. As an example of a parallel connection and the method of determining the nature of the shunt impedance, consider the coaxial-line T junction shown as part of the junction-coupled coaxial-line resonator in Fig. 42A. The T junction may be thought of as equivalent to the parallel connection of Fig. 41B, the device Z being replaced by the input impedance of a coaxial line. An equivalent circuit for the T junction appears in Fig. 42B; the inductance of the connecting lead (PQ in Fig. 42A) is indicated as L_l.

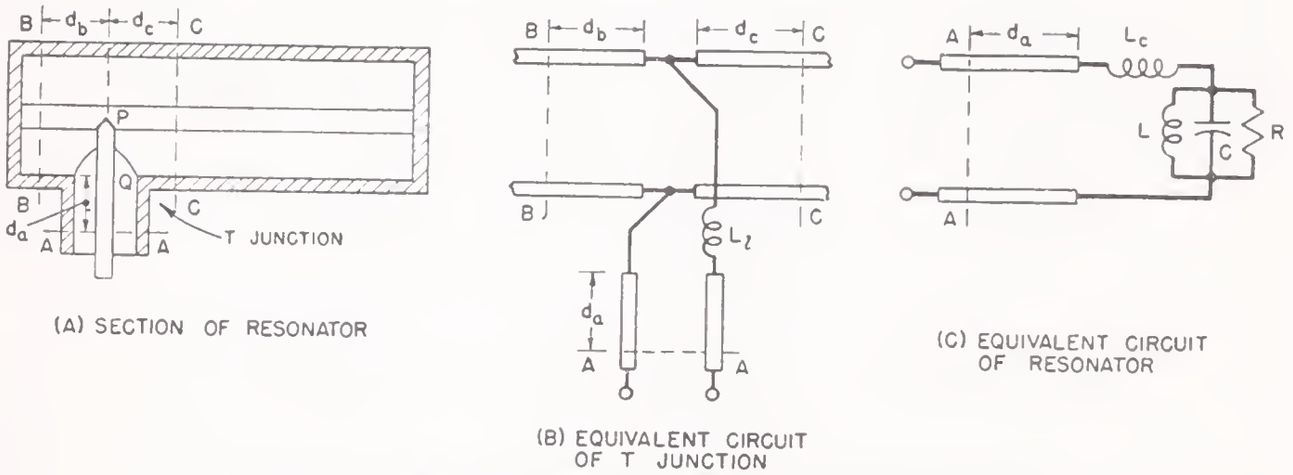


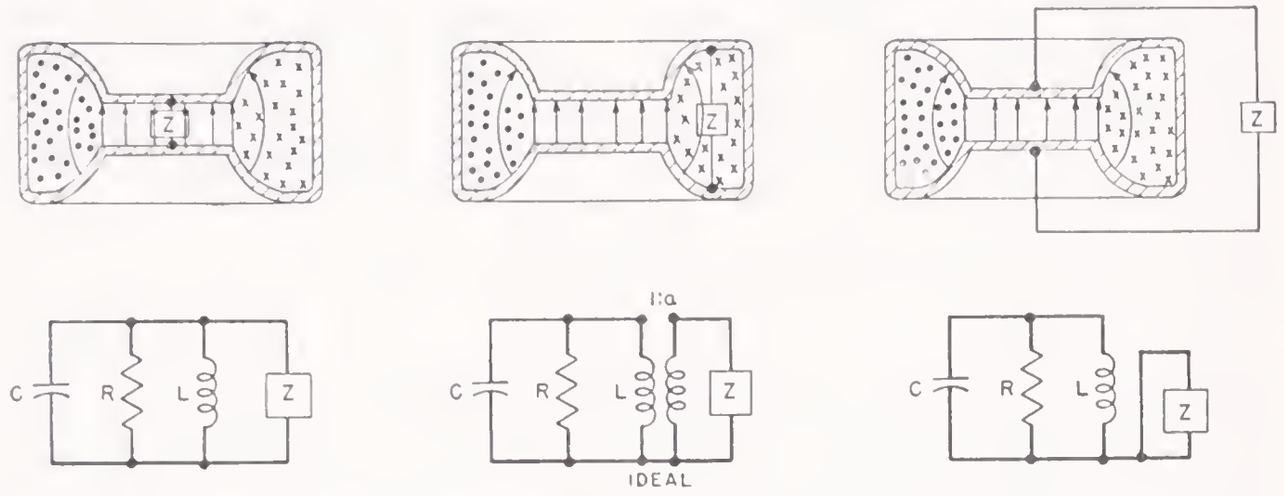
FIG. 42. Coaxial-line resonator with junction coupling.

The justification for the equivalent circuit may be indicated in more detail. In the immediate region of the T junction a complicated field pattern made up of the principal wave plus higher order waves exists. The higher order waves attenuate rapidly with distance from the junction, and reference planes (indicated by the lines AA, BB, and CC) may be chosen so that only the principal wave appears in the regions beyond the reference planes. The equivalent circuit in Fig. 42B thus contains three transmission lines joined by a network that represents the region of higher order waves.

The form of this network may be determined at relatively low frequencies for which the transverse line dimensions are small fractions of a wavelength. The side-arm line may, for these frequencies, be considered to lie in a transverse plane of the main line. A parallel connection is thus indicated. The nature of the parallel impedance is determined by first imagining the side-arm line to be short-circuited at the entrance to the main line by a metal plug. The impedance in shunt with the main line is then the inductance L_l of the post formed by the end PQ of the center conductor. If the short-circuiting plug is removed, the input impedance of the side-arm line appears in series with this inductance, as in Fig. 42B.

At very low frequencies, L_l may be ignored because of its small reactance, and the exact location of reference planes is unimportant because of the great length of low-frequency waves. At higher frequencies, the reactance of L_l is appreciable, and the distances d_a, d_b, and d_c are important in terms of a wavelength. At very high frequencies, a more complicated equivalent circuit is needed. Fortunately, however, the circuit of Fig. 42*B* is moderately accurate throughout the range of frequencies normally employed, this range being limited to frequencies for which only the principal wave is propagated.

The equivalent circuit for the complete resonator of Fig. 42*A* is drawn in Fig. 42*C*. The input line and part of the coupling inductance L_c in this



(A) IMPEDANCE Z LOCATED BETWEEN PLATES

(B) IMPEDANCE Z LOCATED IN TOROIDAL SPACE

(C) IMPEDANCE Z CONNECTED TO OUTER SURFACE

FIG. 43. Sectional views and equivalent circuits for re-entrant cavity with parallel impedances.

circuit correspond to the side-arm line and inductance L_l in Fig. 42*B*. The remainder of L_c and the parallel resonant circuit form the equivalent circuit for a line section with connection to taps on the conductors (see Art. 9).

A second parallel connection of importance is illustrated in Fig. 43, where an impedance element is connected between two points of a re-entrant cavity. As explained in Art. 5, if the electric and magnetic fields occupy different regions in the cavity, the equivalent-circuit representation of the cavity is a parallel *R-L-C* circuit, as indicated in Fig. 43, the inductance being determined by the cross-sectional area and circumference of the toroid and the capacitance being determined by the spacing and area of the parallel plates. The voltage across the parallel combination is the voltage between plates measured along a path perpendicular to the plates in the gap.

Several possible locations for an impedance Z in the cavity are illustrated in Fig. 43. For the location of Z in Fig. 43*A*, the gap voltage is also the voltage across Z, and in the equivalent circuit Z appears in parallel with R, L, and C. In Fig. 43*B*, however, the voltage along the

path of Z is less than the gap voltage because some of the magnetic field fails to link Z and its leads. The impedance of the resonator circuit viewed from the terminals of Z is therefore reduced exactly as it would be if Z were connected to taps on the conductors of a resonant-line section rather than at the maximum-voltage point. The effect of the reduced voltage at Z is indicated by the ideal transformer in the equivalent circuit of Fig. 43*B*, the transformer ratio a being the ratio of the voltage across Z to the voltage between capacitor plates in the resonator. In Fig. 43*C*, the electric field along the path of Z is zero everywhere, and as indicated in the equivalent circuit both ends of Z are at the same potentials—there is no connection between Z and the oscillations of the cavity.

The location of Z within the resonator gap, as in Fig. 43*A*, is encountered in two very important devices. The first is the klystron (see Chap. XI), and in it Z is the impedance of a small section of an electron beam that passes through perforations in the capacitor plates of the resonator. The conductance component of the admittance of the beam can be made negative, with the result that power is supplied from the electrons to the resonator to sustain oscillations in the cavity. The second device is the t-r tube discussed in Art. 11 of Chap. XII. In the t-r tube Z is the impedance of a gas-filled gap between the capacitor plates of the resonator. For small voltages between the plates, Z is essentially an open circuit. If a strong oscillation is initiated in the cavity, however, the gas ionizes, and a discharge across the gap causes the impedance Z to be very low.

A re-entrant resonator connection analogous to the coaxial-line connection of Fig. 41*B* can be used to obtain the equivalent circuit of Fig. 43*A* when the impedance Z must be external to the resonator. Such a connection is seldom used, however, because of the very large impedance of the resonator thus obtained. To reduce the impedance level, the conductor that enters the resonator is ordinarily bent into a small loop that links only a small part of the magnetic flux of the resonator. The coupling loop that results is the subject of the next article.

14. Coupling Loops. Most common among the methods of connecting a coaxial line to a resonant cavity is the use of a small loop formed in the end of the center conductor of the line and inserted into the side of the cavity. An example of loop coupling to a re-entrant cavity is given in Fig. 38*B*, p. 668, and additional examples of loop coupling appear in Fig. 44. A low-loss circular-cylinder cavity is shown in Fig. 44*A*, and the use of two loops in a two-entry coaxial-line resonator is represented in Fig. 44*B*. Observe that each of the loops in the examples is linked by the magnetic flux of the natural mode of interest. Because of these flux linkages, a voltage is generated in the loop by the cavity oscillations, and the loop couples to the mode.

Consider first a one-entry resonator employing a coupling loop. An equivalent circuit is shown in Fig. 45. In this circuit, d is the distance from an arbitrary reference plane (AA in Fig. 44A) to the inner surface of the cavity wall, and the lumped-parameter network represents the impedance at the input terminals (BC in Fig. 44A) of the loop. This impedance may be thought of as a series combination of the low-frequency inductance of the loop and a parallel resonant circuit that represents

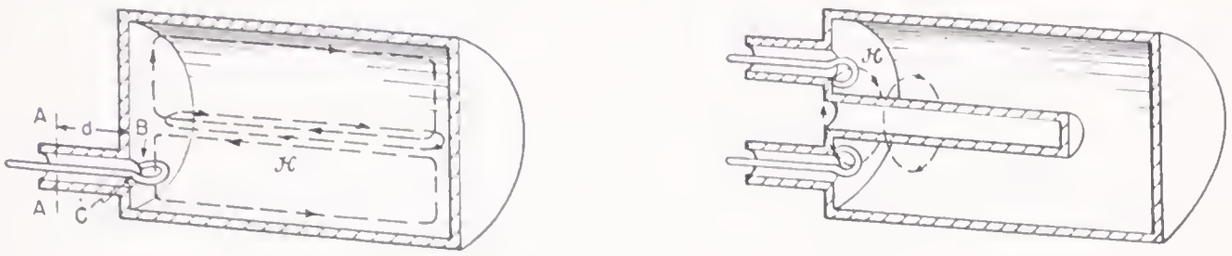


FIG. 44. Cutaway views of cavity resonators with coupling loops.

the effect of the voltage induced by the magnetic field of the cavity oscillations.

The form of the equivalent circuit in Fig. 45 depends upon the relation of the open-circuit and short-circuit natural frequencies of the cavity and coupling-loop combination. If the terminals of the loop are short-circuited by a metallic disk at the cavity end of the coaxial line, a natural frequency slightly higher than f_0 (the natural frequency of the cavity without coupling loop) is obtained. The short-circuited coupling loop is similar to a short-circuited turn on an inductance coil; current in the loop almost completely prevents magnetic flux from linking the loop. Therefore the loop, like the tuning screw and paddle of Art. 5, reduces the region occupied by magnetic flux. The effect is analogous to a reduction of inductance in a lumped-parameter circuit and results in an increase of frequency.

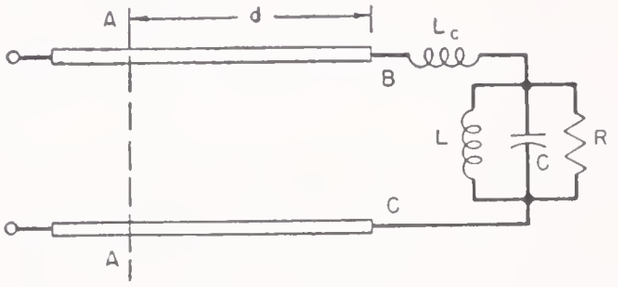


FIG. 45. Equivalent circuit of one-entry resonator employing coupling loop.

If the terminals of the loop are open-circuited—the open circuit may be provided by a short circuit in the input line at a distance of a quarter wave from the loop—the natural frequency obtained is slightly less than f_0. The magnetic-field pattern of the cavity is essentially unchanged by the open-circuited loop, but a local electric field is produced by the voltage between loop terminals. The additional electric field is analogous to a small added capacitor in a lumped-parameter circuit and results in a decrease of the resonant frequency.

Because the short-circuit natural frequency is greater than the open-circuit natural frequency, and because the effect of the loop on the natural frequencies is very small, the input-impedance curve for the loop has a closely spaced peak-dip pair, the dip occurring at a higher frequency than the peak. As explained in Art. 7, an equivalent circuit with coupling inductance is required to represent such an impedance curve, and the general form of the circuit of Fig. 45 is justified. Note that, as may be expected, the coupling loop, like the transformers used at lower frequencies (see Arts. 7 and 9), introduces inductive coupling reactance. For some resonators, the analogy between the coupling loop and a transformer is very close; for others, a secondary winding to be associated with the loop considered as a primary winding is far from evident.

The parameter d in Fig. 45 has already been defined. The other parameters may be determined by the same methods employed for lumped-parameter and transmission-line resonators. As explained in Art. 7, the coupling inductance L_c is approximately the low-frequency inductance of the loop. The resonant-circuit parameters depend upon the resonant frequency, quality factor, and stored energy of the open-circuit natural mode. Because of the very small effect of the loop upon the cavity, little error is incurred by employing the quantities f_0, Q, W, and $\mathcal{H}_m$, which apply to the cavity without the coupling loop. The quantity $\mathcal{H}_m$ is the peak value of the magnetic-field intensity at the location of the loop when the stored energy of the oscillation is W. For a small loop, $\mathcal{H}_m$ may be considered uniform over the area of the loop, and thus the flux linking a loop of area A is

$$\Phi_m = \mu \mathcal{H}_m A \cos \theta \sin \omega t$$

where θ is the angle between the direction of the magnetic flux and a perpendicular to the plane of the loop. The voltage induced in the loop is the derivative of Φ with respect to time and has the peak value

$$E_m = \omega \mu \mathcal{H}_m A \cos \theta \tag{45}$$

From the values of f_0 and Q and the relation of E_m to W, the resonant-circuit parameters may be found as explained in Art. 6. If desired, the circuit of Fig. 45 may be simplified by changing the coupling inductance into an equivalent length of transmission line (see Art. 10). Often the associated changes of impedance level and resonant frequency of the resonant circuit are unimportant.

Coupling loops are often made adjustable so that the impedance level of the resonator and loaded Q of the over-all system can be changed. The adjustment can be made either by partly withdrawing the loop from the cavity, so that the area A in Eq. 45 is changed, or by rotating the loop so that the angle θ is altered.

If loop coupling is used at each entry of a two-entry resonator, the equivalent circuit takes the form of Fig. 46. Inductances L_{c1} and L_{c2} represent the coupling inductances determined at each entry, and d_1 and d_2 are the lengths of the input and output lines measured to the reference plane of each point of entry. For the mode of interest, the voltage induced in one loop is proportional to $\mu\mathfrak{H}_{m1}A_1 \cos\theta_1$, and in the other loop to $\mu\mathfrak{H}_{m2}A_2 \cos\theta_2$, where the symbols have the same meaning as for the one-entry resonator, numerical subscripts denoting the entry. The transformer ratio in the equivalent circuit is thus

$$a = \frac{\mathfrak{H}_{m1}A_1 \cos\theta_1}{\mathfrak{H}_{m2}A_2 \cos\theta_2}$$

In order to determine a for a resonator, it is necessary to know only the sizes and orientations of the loops and the intensities of the magnetic fields at the loop for the mode of interest. In a symmetrical resonator like that of Fig. 44*B*, $\mathfrak{H}_{m1} = \mathfrak{H}_{m2}$ and a depends only upon loop geometry.

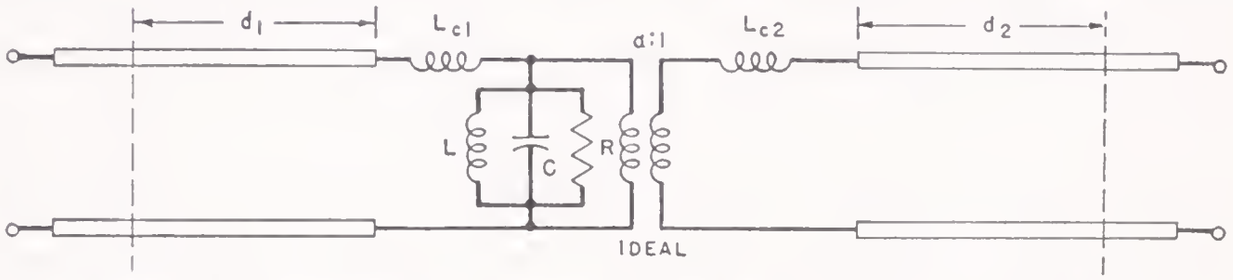


FIG. 46. Equivalent circuit of two-entry resonator employing coupling loops.

15. Wavemeters. Specially designed, tunable cavity resonators called *wavemeters* are used for the measurement of wavelength and frequency of uhf and microwave signals. Wavemeter resonators are usually connected by coupling loops to coaxial lines, and thus they are of interest as applications of the principles developed in the preceding articles.

Cavity-resonator wavemeters may be of either the one-entry or the two-entry type, depending upon the number of coaxial lines coupled to the resonator. The operation of a one-entry wavemeter depends upon a sharp variation of the input impedance that occurs as the cavity is tuned through resonance. The impedance variation is indicated by a corresponding variation of input voltage, and the tuning control is usually adjusted to the point of *minimum* input voltage. The two-entry wavemeter depends upon the variation with tuning of the amount of signal transferred from the input to the output circuit. The output signal is appreciable only if the resonant frequency of the wavemeter is near the signal frequency. The tuning control is adjusted for *maximum* output voltage. The one-entry wavemeter is often referred to as an *absorption-type wavemeter*, and the two-entry wavemeter is called a *transmission-type wavemeter*.

Either type of wavemeter may be used in two ways—either one or two settings of the tuning control may be made in a determination of wavelength. The one-setting method can be used only if an accurate calibration of the wavemeter has been made previously. The tuning control is adjusted for minimum input voltage (or maximum output voltage), and wavelength or frequency is read directly from tuning-control dial calibrations or determined by reference to a calibration chart. The need for previous calibration can be avoided, if the resonator is of the coaxial-line form, by use of the two-setting method. Successive positions of minimum or maximum voltage occur for half-wave changes in the line length.

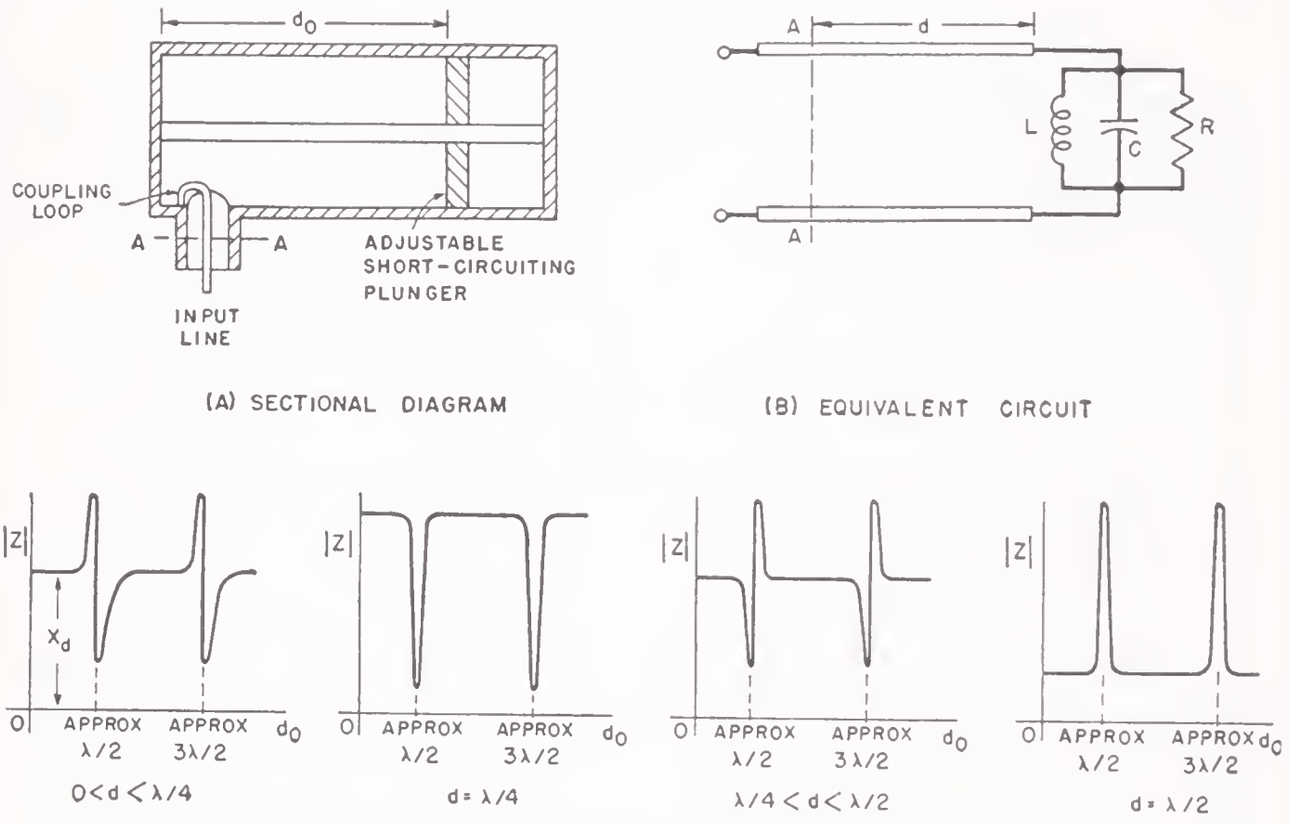


FIG. 47. One-entry coaxial-line wavemeter.

Thus wavelength can be calculated from the difference in the line lengths corresponding to two settings of the tuning control.

One-entry Wavemeter. Figure 47*A* shows the essential features of a one-entry coaxial-line wavemeter. A sinusoidal voltage of unknown frequency is applied to the input line, and the movable plunger is adjusted until a minimum voltage is obtained at the plane *AA* in the input line. The voltage-indicating device at *AA* is not shown in the figure, but it may consist of a crystal rectifier (see Art. 18, Chap. XII) used with a d-c meter. The input impedance at *AA*, and consequently the voltage observed, is essentially independent of the position of the adjustable plunger, except that a wide variation is obtained as the plunger passes through positions that make d_0 in Fig. 47*A* approximate a multiple of a half wavelength.

The manner in which the voltage and input impedance vary as a function of d_0 may be determined from the equivalent circuit in Fig. 47*B*. This circuit is obtained from Fig. 45 if the coupling inductance L_c is replaced by an equivalent length of transmission line, appropriate modifications of the resonant-circuit parameters being made. Distance d in Fig. 47*B* is the sum of the length of the line section that replaces L_c and the distance between plane AA and the inner wall of the wavemeter in Fig. 47*A*. Elements L and C resonate approximately, but not exactly, at a frequency for which d_0 is a multiple of a half wavelength.

The curves of Fig. 47*C* describe the variations of $|Z|$, the magnitude of the input impedance of the circuit of Fig. 47*B*. These curves are similar to those drawn in Fig. 32*B*, p. 659, except that the impedance is plotted for a fixed frequency and variable line length rather than vice versa. For example, if d is less than a quarter wave, as in the left-hand part of Fig. 47*C*, most values of d_0 give rise to a value of $|Z|$ corresponding to X_d, the reactance of a short-circuited line less than a quarter-wave in length. Only for values of d_0 approximating a multiple of $\lambda/2$ is the input line terminated in an impedance appreciably different from a short circuit. If d_0 is slightly less than $\lambda/2$, an open-circuit mode corresponding to resonance of L and C with the capacitance of the input line occurs. If d_0 is slightly more than $\lambda/2$, a short-circuit mode corresponding to resonance of L and C with the inductance of the input line occurs. Thus the peak-dip pair in the $|Z|$ curve is obtained, as shown. For $d = \lambda/4$, $|Z|$ has the high value corresponding to the input impedance of a shorted-end quarter-wave line except for the dips at the half-wave values of d_0, and for $d = \lambda/2$, $|Z|$ has a low value except for sharp peaks at the half-wave points.

The curves of Fig. 47*C* indicate that, if the wavemeter is to be adjusted for minimum input impedance, the most desirable value of d is a quarter wavelength. In wavemeters intended for two-setting operation, d is often made adjustable so that it can be set correctly for the band of frequencies in which the wavemeter is to be employed. Changing the value of d does not impair the accuracy of the results, provided the same value of d is used for both settings in a particular measurement of wavelength. For single-setting wavemeters, however, d must be fixed. Otherwise the calibration of the wavemeter would be changed, because the frequency at which the minimum impedance occurs depends slightly upon the value of d. Use of the single-reading wavemeter is therefore limited to a relatively narrow frequency band in which d is sufficiently close to a quarter wavelength to provide a prominent dip in the $|Z|$ curve.

For precise measurement of wavelength, the loaded Q of the system should be as high as possible in order that the dips may be very sharp. As explained in Art. 14, the loaded Q may be varied by changing the size

of the coupling loop and its orientation with respect to the magnetic field. Designing the loop for small flux linkage yields a low-impedance resonant system and consequently a high loaded Q. The flux linkage cannot be made too small, however, because for extremely small flux linkages the dips are of very low amplitude and difficult to locate. Final loop size therefore represents a compromise between ease of operation and precision of measurement.

Because of the variations of the resonator input impedance that occur as the short-circuiting plunger is adjusted, the impedance presented to the source at the input terminals of the coaxial line also varies over a wide range. These variations may cause the source frequency to shift as the

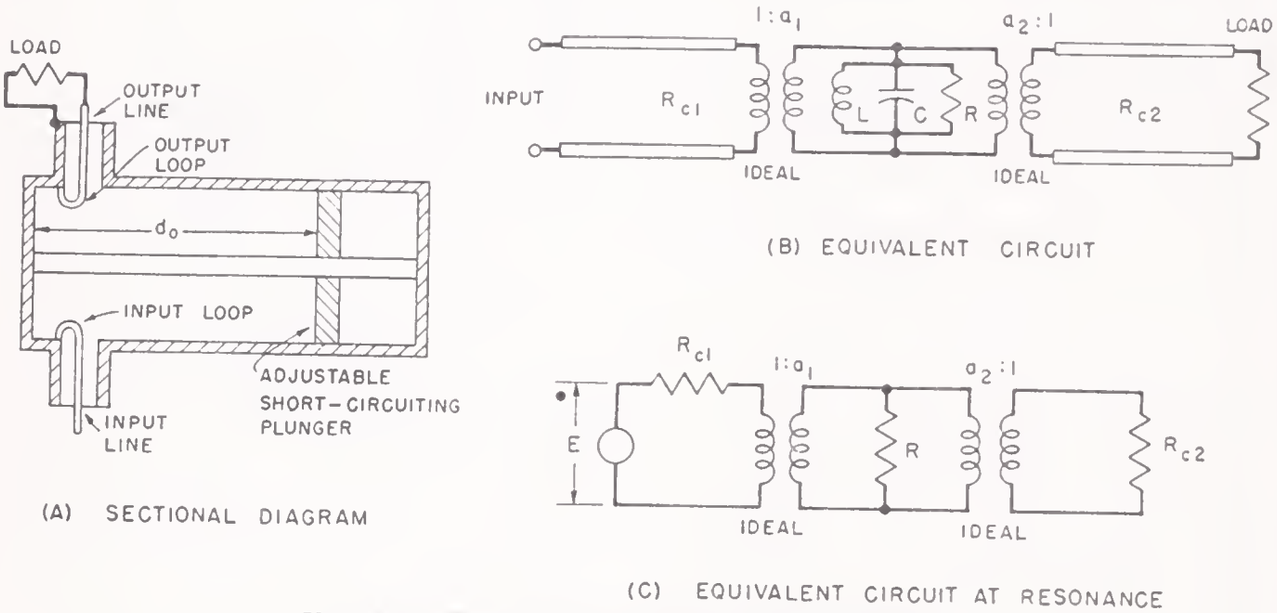


FIG. 48. Two-entry coaxial-line wavemeter.

wavemeter is tuned. To avoid such an effect, it is customary to introduce attenuation between the source and wavemeter. The attenuation may be in the form of a high-loss connecting cable or an attenuator that presents to the lines at each end of the attenuator a resistance equal to their characteristic resistance. Either method prevents any appreciable portion of a wave reflected at the wavemeter from reaching the source and makes it possible to consider that the wavemeter is connected to a source of internal impedance equal to the characteristic resistance of the line. The voltage at plane AA in Fig. 47A therefore varies in approximately the same manner as the impedance at the same plane.

Two-entry Wavemeter. A sectional sketch of a coaxial-line form of two-entry wavemeter appears in Fig. 48A. A signal of unknown frequency is applied to the input line, and the movable plunger in the resonator is adjusted until a maximum voltage is obtained at the load. The wavelength or frequency of the signal may be determined from the setting obtained if the resonator has been calibrated previously. Alternatively, two settings may be made and the wavelength calculated from the differ-

ence in line lengths obtained. To prevent reaction of the wavemeter on the source, a high-loss input line or an attenuator in the input circuit is required as with the one-entry wavemeter.

Details of the operation of a two-entry wavemeter may be explained in terms of the equivalent circuit of Fig. 48*B*, which is obtained in the same manner as that of Fig. 47*B*. The resonant frequency of the *R-L-C* circuit is very near one of the frequencies for which d_0 is a multiple of a half wavelength, and the ideal-transformer ratios depend upon the flux linkages of the input and output coupling loops, as explained in Art. 14. Because the loops are small, the loaded Q of the system is high, the tuned circuit is effectively a short circuit except at frequencies near resonance, and appreciable power reaches the load only when the wavemeter is tuned to resonance.

The coupling loops must be small in order that the tuned circuit may be an effective short circuit at frequencies very close to resonance and thus provide sharp tuning of the wavemeter. However, if the loops are made too small, the tuned circuit is very nearly a short circuit even at resonance and the load voltage may become too small to detect. The choice of the loop size is therefore a compromise between the requirements of selectivity and ease of detecting the output.

Not only can both coupling loops be made large or both small, as assumed in the preceding paragraph, but also one loop may be made large or small relative to the other. Assume, for simplicity, that the load is a resistance equal to R_{c2}, the characteristic resistance of the output line, and that the input line is connected to a source of internal impedance equal to R_{c1}, the characteristic resistance of the input line. The equivalent circuit at resonance then takes the form indicated in Fig. 48*C*. If the conductance of R is ignored, maximum power is delivered to R_{c2} when the loops are adjusted so that $(a_2/a_1)^2 R_{c2}$ is equal to R_{c1}. (See the maximum-power condition in Art. 13, Chap. VI.) If $R_{c1} = R_{c2}$, the flux linkages of the two loops should be equal.

The coaxial-line form of wavemeter resonator illustrated in Fig. 48 is, of course, required if measurements are to be made by the two-setting procedure. On the other hand, other forms of cavities may be employed if a calibration curve and the one-setting procedure are used. Particularly useful for microwave measurements is the circular-cylinder cavity operated in a low-loss $TE_{0,1,n}$ mode. Because of its very high Q, it is very sharply selective and can be made to show as separate response peaks the output signals arising from relatively closely spaced Fourier components of complex waveforms. When used in this manner with a complex waveform, a wavemeter is called a *spectrum analyzer*, and the ability to separate wave components closely spaced in frequency is called *resolving power*.

The possibility of use with a complex waveform is a special advantage

of the two-entry wavemeter. A one-entry wavemeter is useful only with essentially sinusoidal input waves—the large off-resonance response obtained from one component of a complex wave would obscure the minimum in the response produced by another component. If the input wave is sinusoidal, however, the one-entry resonator has an advantage in convenience because the off-resonance response gives indication that the equipment is functioning.

16. Coupling Probes. An alternate to the coupling loop that is particularly useful if the coaxial-line center conductor must be insulated from the resonator wall for low-frequency signals is the *coupling probe*. Unlike the loop, the probe does not provide mechanical support for the end of the center conductor, and an insulating bead may be required at

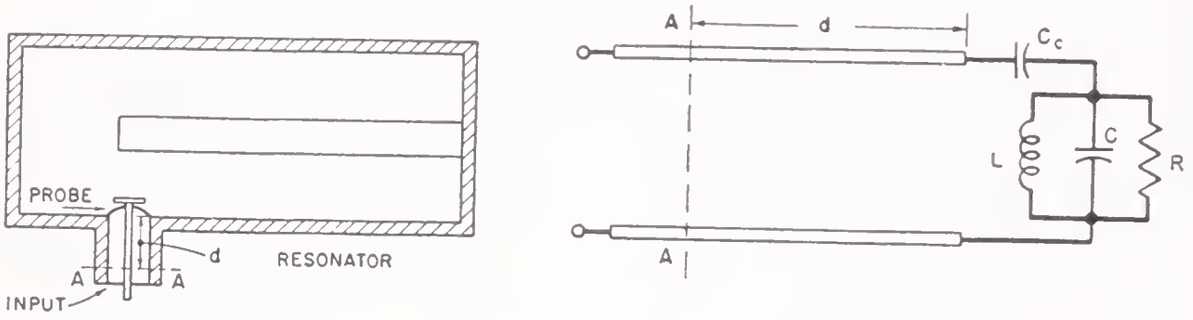


FIG. 49. Coaxial-line resonator with probe coupling.

the entrance to the cavity. One example of a probe in a cavity resonator is sketched in Fig. 38*C*, p. 668. The probe in this figure is simply an extension of the center conductor of the coaxial line. Such a probe is frequently employed, but often, also, the end of the probe is enlarged; for example, a disk may be affixed to the end of the probe as in Fig. 49*A*.

The equivalent circuit for a resonator with a coupling probe may be obtained by the same general method employed with the coupling loop. With a short circuit at the cavity end of the coaxial line, the resonant frequency of the probe and cavity is somewhat less than the natural frequency of the cavity. This shift of frequency is apparent for the resonator of Fig. 49*A*, because the probe constitutes a capacitance at the end of the resonator line. When the probe terminals are open-circuited, the reduction of the resonant frequency caused by the probe is decreased; the open-circuited probe may be considered equivalent to two capacitors in series. Thus the short-circuit resonance occurs at a lower frequency than the open-circuit resonance, and the peak-dip pair in the input-impedance curve is represented by an equivalent circuit with coupling capacitance, as in Fig. 49*B*.

The coupling capacitance, C_c in Fig. 49*B*, is approximated by the low-frequency capacitance between the probe and the resonator conductor. The other parameters in Fig. 49*B* can be calculated by the methods of

Art. 6, provided the frequency, Q, and relation of probe voltage to stored energy for the open-circuit mode are known. The method of determining the probe voltage for a given stored energy can be explained by means of a different and more detailed derivation of the equivalent circuit for the resonator of Fig. 49*A*.

This second derivation makes use of Fig. 50. In Fig. 50*A* is shown the resonator of Fig. 49*A* with the input line removed but with the disk from the end of the probe still in place. The disk has little effect upon the field configuration in the cavity because it is thin in the direction parallel to the electric-field lines. The section of the resonator to the right of plane *BB* can therefore be represented, as in Fig. 50*B*, by the line section

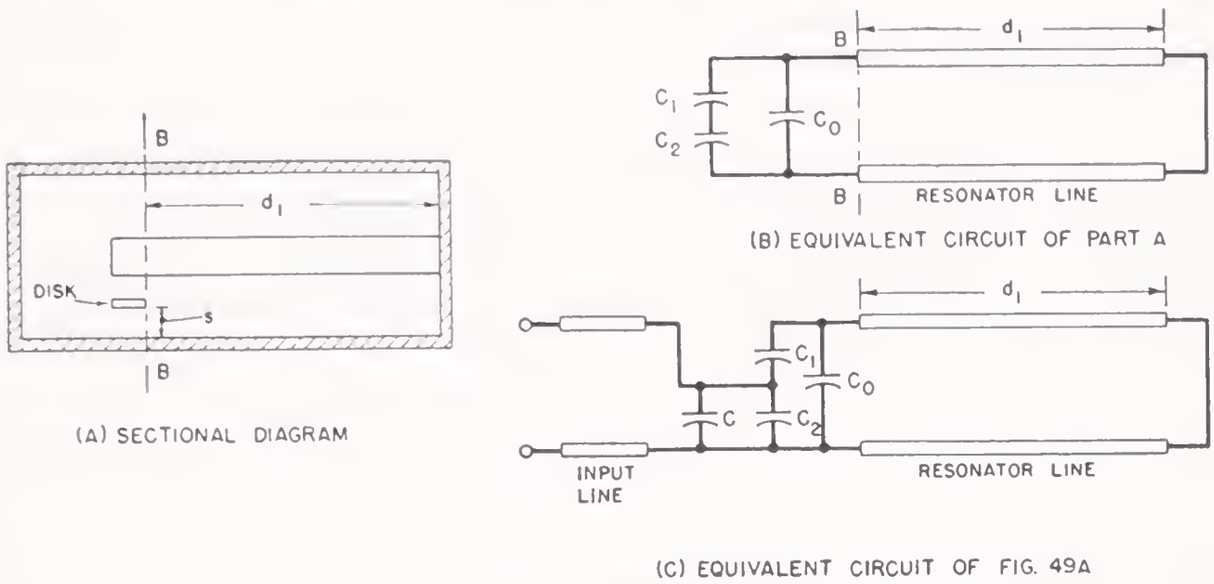


FIG. 50. Resonator and equivalent circuits used in explanation of probe coupling.

of length d_1. As indicated in the figure, the part of the resonator to the left of *BB* is equivalent to a network of three capacitances: C_0 representing the capacitance between center conductor and outer wall, C_1 representing the capacitance between center conductor and disk, and C_2 representing the capacitance between disk and outer conductor. As explained in Art. 13, an impedance located between the disk and the outer conductor would appear in parallel with C_2 in the equivalent circuit. The input line of Fig. 49*A* provides a means of making such a connection, and thus the complete equivalent circuit of Fig. 50*C* is obtained. The inductance corresponding to L_1 in Fig. 42*B*, p. 672, is omitted because the length of probes is ordinarily very short. On the other hand, the capacitance between the probe rod and the outer wall, which may be comparable to C_2, is represented by C in Fig. 50*C*.

The close resemblance of the resonator and equivalent circuit of Figs. 49*A* and 50*C* to the resonator and equivalent circuit of Fig. 31, p. 657, representing capacitive coupling to a parallel-wire line should be noted. The circuit of Fig. 50*C* can be reduced to the one in Fig. 49*B* by the same methods employed in connection with Fig. 31.

The probe voltage corresponding to a given stored energy can be calculated from Fig. 50*A*, provided the characteristic resistance of the line and the values of the capacitances in the figure are known. Were it not for capacitor C, the probe voltage would be equal to the voltage between the disk and the outer conductor in Fig. 50*A*. Thus a first approximation to the probe voltage is $\mathcal{E}s$, where $\mathcal{E}$ is the electric-field intensity (measured with the probe removed) in the region occupied by the probe for a given stored energy and s is the distance indicated in Fig. 50*A*. The probe is assumed to be short, so that $\mathcal{E}$ may be considered uniform over the probe length.

Corresponding to the general principle that the open-circuit voltage of a coupling loop is the voltage induced by the magnetic flux linking the loop is the following principle for a probe: The open-circuit voltage of a probe is equal to the voltage $\mathcal{E}s_{\text{eff}}$ short-circuited by the probe, s_{eff} being the *effective probe length*. The value of s_{eff} varies in relation to the over-all probe length s because of the effect of capacitance C in Fig. 50*C*. For a probe with a large terminating disk, s_{eff} is only slightly less than s; for a probe that is a straight extension of the coaxial-line center conductor, s_{eff} is of the order of half the total probe length.

The ratio of s_{eff} to s is related to the current distribution in the probe. When a large terminating disk is used, the current is essentially uniform throughout the probe length and equal to the current received by the probe from the coaxial line. If the disk is omitted, the current decreases to zero at the tip of the probe because of the distribution of capacitance along the probe length. Thus the end of the probe is not fully effective in coupling the probe to the resonator, and the effective probe length is reduced.

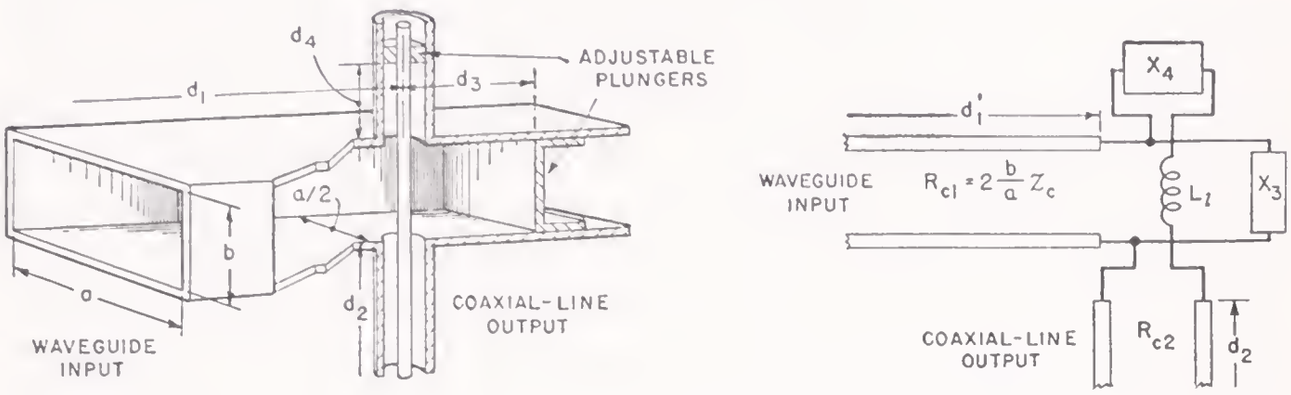
The advantage of the effective-length concept is that it applies not only to the particular resonator of Fig. 49*A* but to a probe in any cavity resonator. For example, if a probe penetrates a rectangular-waveguide cavity in the direction of the electric-field lines, the capacitances in Fig. 50*C* are not easily specified. Nevertheless, an equivalent circuit like Fig. 49*B* can be obtained, the impedance level being calculated from the relation of $\mathcal{E}s_{\text{eff}}$ to the stored energy.

The probe, like the loop, provides a simple means of adjusting the loaded Q of the resonator-line system. Increasing the penetration of the probe increases the effective length and (like an increase of the area of a loop) increases the coupling and decreases the loaded Q. Usually a probe is inserted in a direction parallel to the electric-field lines of the mode of interest. Should the probe be inclined at an angle θ to the field, the probe voltage is reduced to $\mathcal{E}s_{\text{eff}} \cos \theta$.

17. Waveguide-to-coaxial-line Junctions. As explained in Art. 12, cavity resonators may be connected to waveguides by means of an inter-

mediate section of coaxial line. A direct junction, a loop, or a probe may serve to couple the resonator to the line, and the line-to-guide junction may also be based on any of these three coupling devices.

Figure 51*A* shows one form of a junction-type connection between a coaxial line and a rectangular waveguide. The guide, if operated so that only $TE_{1,0}$ waves are propagated, may be considered analogous to a parallel-strip transmission line (see Art. 7, Chap. VIII), and the junction may be regarded as an impedance shunting the guide. The nature of the shunt impedance may be determined by the same method employed with the coaxial-line T junction (see Art. 13). Assume that the two coaxial lines in Fig. 51*A* are temporarily short-circuited at the top and bottom walls of the guide by metallic plugs. The conductor extending



(A) CUTAWAY VIEW (B) EQUIVALENT CIRCUIT

FIG. 51. Junction coupling between waveguide and coaxial line.

through the guide is then an inductive post (Art. 14, Chap. VIII) and may be represented in an equivalent circuit by an inductance L_l. Removing the short-circuiting plugs places the two lines in series with this inductance, and thus the circuit of Fig. 51*B* is obtained. In this circuit, X_3 represents the input reactance of the short-circuited guide section of length d_3 and X_4 represents the input reactance of the short-circuited line section of length d_4. Adjusting the plungers in Fig. 51*A* changes the distances d_3 and d_4 and alters the values of the reactances X_3 and X_4.

In Art. 13 of Chap. VIII, a waveguide is represented in an equivalent circuit by a transmission line of characteristic resistance equal to the characteristic wave impedance of the guide and of the same electrical length as the guide. In Fig. 51*B*, the length d_1' is made less than the length d_1 of the guide itself in order that the electrical lengths may be equal, but the characteristic resistance R_{c1} of the line representing the guide is different from Z_c. The change is necessary because transmission-line symbols for both the guide and the coaxial line appear in the same equivalent circuit. The characteristic resistance R_{c2} of the coaxial line refers to a voltage-current ratio. The characteristic resistance R_{c1} must

also refer to a voltage-current ratio (not to the ratio of $\mathcal{E}$ to $\mathcal{H}$ indicated by $\mathcal{Z}_c$) in order for the circuit to be self-consistent.

The value $2(b/a)\mathcal{Z}_c$ indicated for R_{c1} in Fig. 51*B* may be obtained in the following manner: Suppose that distance d_3 is a quarter guide wavelength, so that X_3 becomes the reactance of an open circuit, and suppose that the adjustment of d_4 yields series resonance of X_4 and the reactance of L_l. Figure 51*B* then indicates a direct connection between the waveguide and the coaxial line. By proper choice of the coaxial-line dimensions, the line can be made to match the guide, so that a wave may travel from the guide into the line with no reflection at the junction. The value of R_{c2} that yields a match when d_3 and d_4 are adjusted in this way is then the value to be used as R_{c1}.

Let $\mathcal{E}_m$ be the peak value of the electric-field intensity at the center of the guide for a wave entering the guide from the left in Fig. 51*A*. The voltage from top to bottom of the waveguide along the center line is $\mathcal{E}_m b$, and this is the voltage applied to the shunt element at the junction. Under the conditions assumed, X_4 and the reactance of L_l are in resonance, and the voltage $\mathcal{E}_m b$ is applied to the coaxial line. If no reflections occur, the power carried to the junction by the wave in the guide is equal to the power carried away by the wave in the line. Because R_{c1} is equal to the value of R_{c2} in this matched condition, the power in the line is $\frac{1}{2}(\mathcal{E}_m b)^2/R_{c1}$. As shown in Art. 9, Chap. VIII, the power in the guide is $\frac{1}{4}\mathcal{E}_m{}^2 ab/\mathcal{Z}_c$. Therefore

$$\frac{1}{4}\frac{\mathcal{E}_m{}^2 ab}{\mathcal{Z}_c} = \frac{1}{2}\frac{\mathcal{E}_m{}^2 b^2}{R_{c1}}$$

or

$$R_{c1} = 2\frac{b}{a}\mathcal{Z}_c \tag{46}$$

as indicated in Fig. 51*B*. This value of R_{c1} should be thought of, not as a property of the guide itself, but as a property of the guide with the particular coupling device in Fig. 51. A coaxial line having this value of characteristic resistance is perfectly matched to the guide by this junction (for the values of d_3 and d_4 assumed) for signals that traverse the junction in either direction.

The adjustments of d_3 and d_4 assumed in the above discussion are not the most useful adjustments for practical purposes. The value of $\mathcal{Z}_c$ is, by Eq. 67 of Chap. VIII, p. 598, $\mathcal{Z}_i/\sqrt{1-(\lambda_0/\lambda_{co})^2}$. The value of $\mathcal{Z}_i$ is 377 ohms, and under normal operating conditions the value of λ_0/λ_{co} is about $\frac{2}{3}$. Thus, by Eq. 46, the value of R_{c1} for a guide for which $b/a = \frac{1}{2}$ is approximately

$$R_{c1} = 2 \times \frac{1}{2} \times 377 \times \frac{1}{\sqrt{1-(\frac{2}{3})^2}} \approx 500 \text{ ohms}$$

Normal coaxial-line dimensions yield values of R_{c2} of 50 to 70 ohms, and therefore a considerable mismatch results if the plunger adjustments assumed above are made.

If the junction is part of a cavity-resonator system, this mismatch is not important. As long as the junction provides some coupling between waveguide and line, and provided the line couples to a resonator, the entire system can be represented by a parallel R-L-C circuit at some reference plane in the waveguide (see Art. 10). Only one adjustable element is required to control the loaded Q of the system, and either of the plungers in the guide-to-line junction or an adjustable element at the line-to-cavity junction can be used to control Q_l.

On the other hand, mismatch is important if a junction of this kind is employed in a transmission system. If power flows from guide to line, an impedance-changing device at the load is adjusted to make the coaxial line flat, and the plungers at the junction are adjusted to eliminate standing waves in the guide. Suppose, for example, that R_{c1} in Fig. 51B is 500 ohms and R_{c2} is 50 ohms, and let X be the reactance of X_4 and L_l in series. The impedance connected across the waveguide is then $R_{c2} + jX$, and the conductance component of the corresponding admittance is $R_{c2}/(R_{c2}^2 + X^2)$. The susceptance component of this admittance can be canceled by proper adjustment of X_3, and the line representing the waveguide in Fig. 51B is then terminated in a resistance of value $(R_{c2}^2 + X^2)/R_{c2}$. For $R_{c2} = 50$ ohms, the desired value of 500 ohms terminating the waveguide can be obtained if X_4 is adjusted so that the magnitude of X is 150 ohms.

Because large values of series reactance and shunt susceptance are required to produce an impedance match in this way, the performance of this junction is very sensitive to frequency, and the match is maintained over only a narrow frequency band. Several other types of junctions are available for use where greater bandwidth is required. One of these, illustrated in Fig. 52A, is based upon the coupling probe and provides sufficient bandwidth for most radar applications.

As explained in Art. 16, the probe may be thought of as a capacitor voltage divider across the waveguide, the center conductor of the coaxial line being connected to the tap point of the divider. If $\mathcal{E}_m$ is the peak electric-field intensity at the center of the guide and s_{eff} is the effective length of the probe, then $\mathcal{E}_m b$ is the voltage across the entire divider and $\mathcal{E}_m s_{\text{eff}}$ is the voltage applied to the coaxial line. The voltage divider can therefore be replaced by an ideal transformer with shunt and series capacitance, the ratio of the transformer being s_{eff}/b (Art. 20, Chap. II). Figure 52B results when this entire circuit is referred to the coaxial-line side of the transformer. The series capacitance is C_c, the shunt capacitance combines in parallel with the reactance of the waveguide section of

length d_3 to form reactance X_3, and the value indicated for R_{c1} is obtained through multiplying the value $2(b/a)Z_c$ (applicable for a connection completely across the guide) by s_{eff}^2/b^2, the square of the transformer ratio.

An impedance match is obtained at this junction if the probe penetration is adjusted so that $1/R_{c1}$ is equal to the conductance component of the admittance of R_{c2} and C_c in series, and the waveguide plunger is adjusted so that $1/X_3$ cancels the susceptance component of this admittance. If the reactance of C_c is small relative to R_{c2}, the series reactance

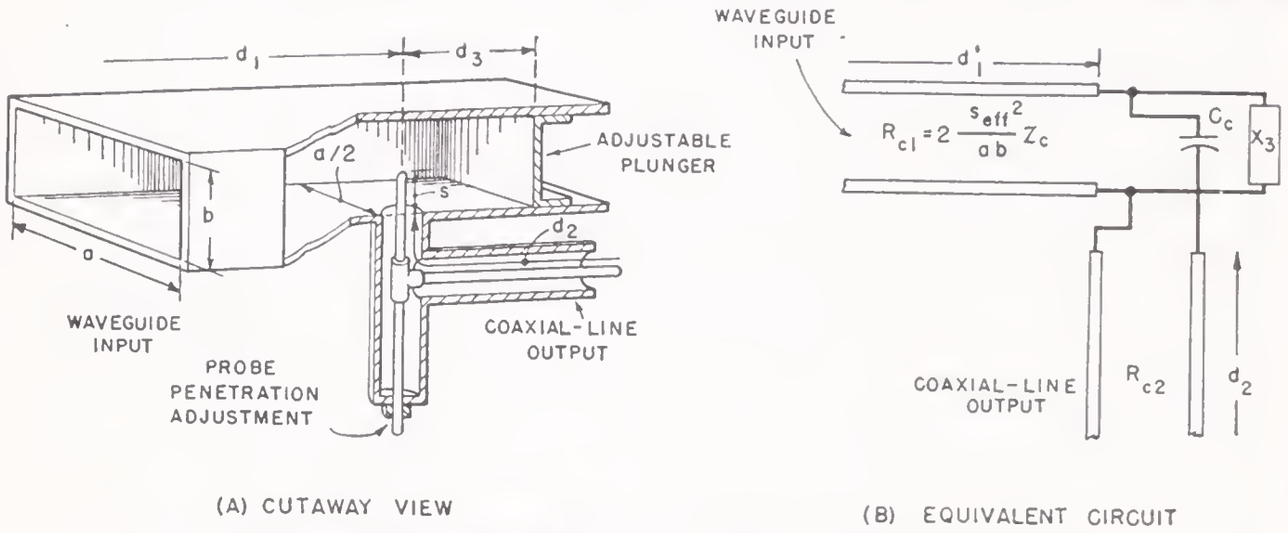


FIG. 52. Waveguide-to-coaxial-line junction employing probe coupling.

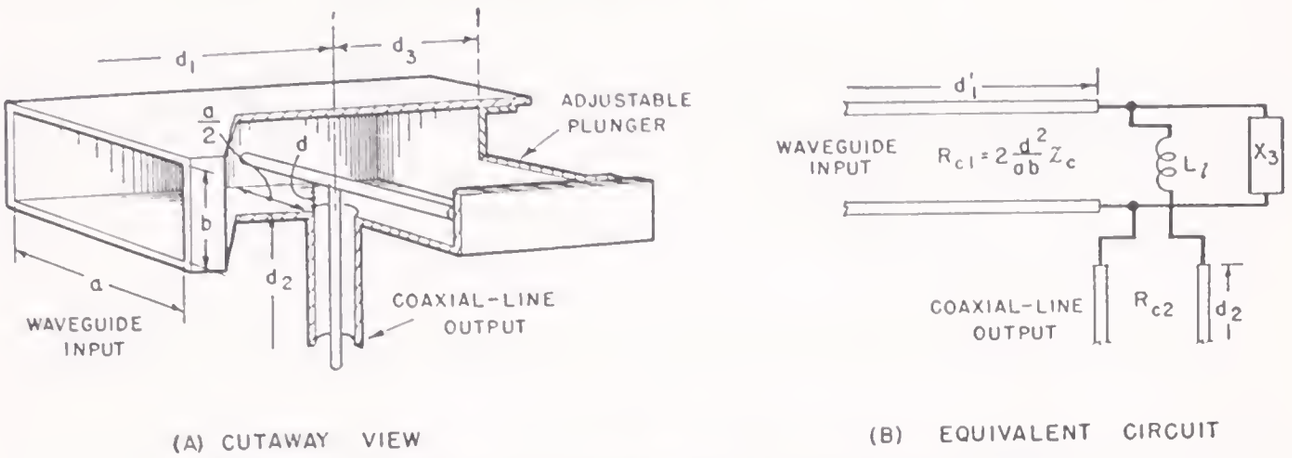
and shunt susceptance introduced by these adjustments are relatively small, and much greater bandwidths are obtainable with this junction than with the junction of Fig. 51*A*.

Suppose that to the end of the probe in Fig. 52*A* is attached a short rod extending parallel to the *a* dimension of the guide. Such a rod, like the terminating disk considered in Art. 16, increases the capacitance C_c and causes the effective probe length s_{eff} to approach the actual length s. If the length of the transverse rod is increased until the ends of the rod touch the side walls of the guide, a third type of guide-to-line junction (illustrated in Fig. 53*A*) is obtained. In the equivalent circuit (Fig. 53*B*), capacitance C_c is omitted because it is short-circuited when the transverse rod touches the guide walls, and inductance L_l is introduced to represent the inductance of the probe and rod. In the relation for R_{c1}, the distance d (defined in Fig. 53*A*) replaces the effective length s_{eff} of the probe of Fig. 52.

Adjustment of the junction of Fig. 53*A* for an impedance match is the same as that of the probe junction, except that distance d must be determined correctly when the device is made and cannot conveniently be adjusted thereafter. Distance d is chosen so that $1/R_{c1}$ is equal to the conductance component of the admittance of R_{c2} in series with L_l, and the plunger is set so that $1/X_3$ cancels the susceptance of this admittance. The series reactance and shunt susceptance thus introduced are not large,

and the bandwidth obtained is adequate for most radar applications.

The junction of Fig. 53*A* may be thought of from an alternate point of view as a form of loop coupling. Each half of the transverse rod together with a return path in the guide walls forms a loop, and these two loops are connected in parallel to the coaxial line. From this point of view, the junction of Fig. 51*A* (with d_4 set equal to zero) is also an example of loop coupling. Because each of the two loops for Fig. 51*A* is linked by the entire magnetic flux between the center conductor and one side of the



(A) CUTAWAY VIEW (B) EQUIVALENT CIRCUIT

FIG. 53. A third form of waveguide-to-coaxial-line junction.

waveguide, a high voltage is applied to the coaxial line. Therefore, the impedance of the waveguide circuit to which the coaxial line connects is undesirably high, as explained in the discussion of Fig. 51. The arrangement of Fig. 53*A* reduces the flux linkage and permits making R_{c1} comparable to R_{c2}. An alternate procedure for securing the same result is to move the coaxial line of Fig. 51*A* from the center toward one side of the guide.

The junctions discussed in this article are only three of a considerable variety of waveguide-to-coaxial-line junctions that have been developed.[1] Some of these provide special advantages of high power capacity or exceptionally great bandwidth.

18. Aperture Coupling. Apertures, such as those illustrated in Fig. 39, p. 669, are commonly used to connect either coaxial lines or waveguides to cavity resonators, especially where adjustable coupling is not required. Apertures are physically simple, but calculation of their performance characteristics is often difficult.

A waveguide-and-cavity system particularly useful in an explanation of aperture coupling is shown in Fig. 54*A*. The rectangular guide is assumed to transmit $TE_{1,0}$ waves only, and the $TE_{1,0,1}$ mode (z axis along the length of the guide) is considered in the cavity. Coupling is provided by an aperture in the metallic plate between guide and cavity. This sys-

[1] See, for example, GEORGE L. RAGAN, *Microwave Transmission Circuits*, 332–339 (McGraw-Hill Book Company, Inc., New York, 1948).

tem is of special interest because it may be thought of as a continuous waveguide closed at one end and containing an iris at point P.

An equivalent circuit for Fig. 54A is drawn in Fig. 54B. In this circuit, losses are ignored, and transmission lines A and B represent the two guide sections. The electrical lengths of the lines are equal to the electrical lengths of the corresponding guide sections, and the half-wave mode of line B corresponds to the $TE_{1,0,1}$ mode of the cavity. Reactance

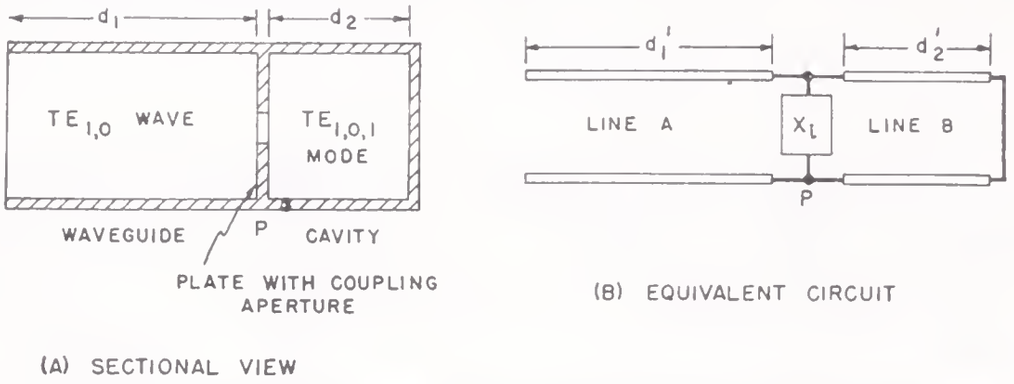


FIG. 54. Rectangular waveguide and cavity with aperture coupling.

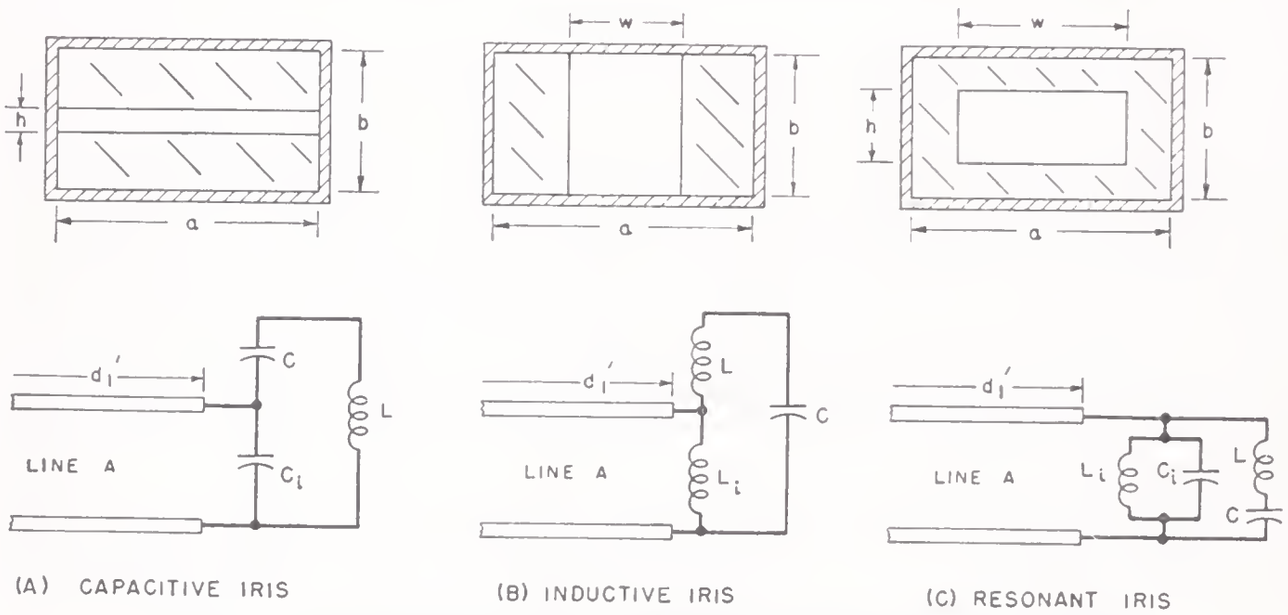


FIG. 55. Equivalent circuits for Fig. 54A for three aperture shapes.

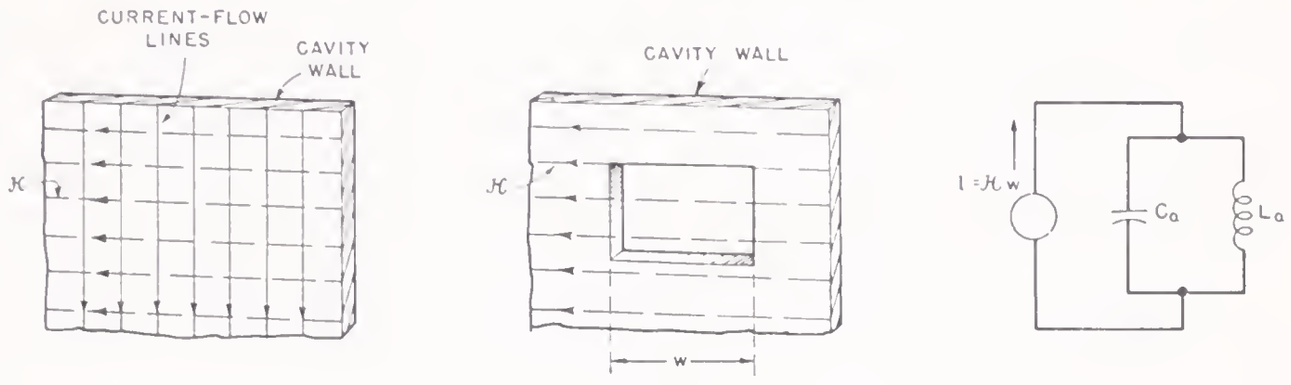
X_i represents the iris and has a value that depends upon the size and shape of the aperture.

If a capacitive iris is employed, the equivalent circuit takes the form indicated in Fig. 55A. If the operating frequency is well above the cutoff frequency of the $TE_{1,0}$ waves, the capacitive reactance X_i is very nearly inversely proportional to frequency (see Eq. 70 of Chap. VIII, p. 606), and representation of the iris by capacitance C_i is correct. The series L-C circuit represents line B and resonates at the natural frequency of the $TE_{1,0,1}$ mode.

Similar equivalent circuits applicable when an inductive iris or a resonant iris is employed are drawn in Figs. 55B and C. If the resonant frequency of the iris of Fig. 55C is equal to the natural frequency of the

cavity, the iris is equivalent within a limited frequency band to a completely open guide. Near the resonant frequency, the admittance of L_i and C_i in parallel is negligible relative to that of L and C in series.

The equivalent circuits of Fig. 55 indicate that a capacitive iris gives rise, as may be expected, to a capacitive coupling reactance for the resonator that terminates line A. Similarly, the inductive iris gives rise to an inductive coupling reactance. The circuits show also that the loaded Q decreases (the coupling is increased) as the aperture is opened, because opening the aperture increases the reactance of C_i or L_i. Note, however, that the shape as well as the size of the aperture has an important effect upon the amount of coupling provided. For example, a resonant aperture of relatively small open area may provide the same coupling as a fully open inductive or capacitive iris. Also, a capacitive



(A) MAGNETIC-FIELD AND CURRENT-FLOW LINES

(B) RECTANGULAR APERTURE IN WALL

(C) EQUIVALENT CIRCUIT OF APERTURE

FIG. 56. Interruption of cavity-wall current by aperture.

iris of small opening provides much more coupling than does an inductive iris of equal open area. This fact may be shown by calculations based upon Eqs. 70 and 71 of Chap. VIII, p. 606, and upon the equivalent circuits of Fig. 55. If the open area of each iris is one tenth the cross-sectional area of the guide, the loaded Q for the inductive iris is approximately 400 times that obtained with the capacitive iris.

Thus far the particular waveguide-cavity system pictured in Fig. 54*A* has been discussed. The principles developed may be applied to apertures in general. Assume that an aperture is to be cut in the cavity wall illustrated in Fig. 56*A*. Natural-mode oscillations in the cavity give rise to a magnetic field of intensity $\mathcal{K}$, and current flows in the wall in a direction perpendicular to the magnetic field as indicated.

Cutting the aperture in the wall interrupts a portion of this current, and the interrupted current may be considered to flow through an impedance formed by the aperture. If the aperture is rectangular and of width w in the direction of $\mathcal{K}$, as in Fig. 56*B*, the interrupted current I is equal to $\mathcal{K}w$ (see Eq. 3 of Chap. VIII, p. 549). Part of this current flows as a displacement current in the capacitance between top and bottom edges

of the aperture, and the remaining part follows a path through the metal around the aperture. Thus in the equivalent circuit of the aperture (Fig. 56*C*), the interrupted current I divides between C_a, the capacitance between top and bottom edges of the aperture, and L_a, the inductance of the path around the aperture. An electric field is associated with the displacement current through C_a, and the perpendicular relation between the electric and magnetic fields in the aperture gives rise to a Poynting vector directed through the aperture. Thus power is transferred through the aperture.

The equivalent circuit of Fig. 56*C* is helpful in an explanation of the effect of aperture dimensions upon the amount of coupling provided by the aperture. Consider a cavity coupled by an aperture to a coaxial line or waveguide, and assume a matched load at the end of the line or guide. For natural-mode oscillations in the cavity of angular frequency ω_0 and stored energy W, let P be the power loss in the cavity and let P_L be the power that flows through the aperture and along the line or guide to the load. By Eq. 6, p. 618, and Art. 10, the loaded Q of the system is $\omega_0 W/(P + P_L)$. Therefore any change in the aperture that increases P decreases the loaded Q and increases the coupling.

An increase in P may result from an increase in aperture area or from an increase in the density of power flow through the aperture. Therefore, the coupling provided by the aperture may be increased either by an increase in the size of the aperture or by any change of aperture shape or location that increases the product of $\mathcal{E}$ and $\mathcal{H}$ (which is the magnitude of the Poynting vector) at points in the aperture. The electric-field intensity depends upon the voltage developed across the aperture reactance by the interrupted current (see Fig. 56*C*). The coupling may be increased, therefore, by locating the aperture at a position of greater magnetic-field intensity, by orienting the aperture with its longest dimension parallel to the magnetic field so that a large current is interrupted, or by adjusting the aperture shape so that the aperture reactance is high. Note that the cavity and transmission system associated with an aperture, as well as the dimensions of the aperture itself, influence the reactance of an aperture. For example, the aperture in the capacitive iris of Fig. 55*A* is represented by the capacitance C_i. If the a dimension of the guide were increased with no change in aperture dimensions, an inductance would be required in parallel with C_i in the equivalent circuit.

The above principles may be applied to the aperture-coupled coaxial-line system of Fig. 39*B*. A sectional view of the system is drawn in Fig. 57*A*, and equivalent circuits appear in Figs. 57*B* and *C*. The outer conductor of the input line may be considered joined to the short circuit representing the end plate by a parallel L-C circuit comprising the capacitance between faces of the aperture and the inductance of the portion

of the conductor containing the aperture. The outer conductor of the resonator line joins the end plate through this same impedance, and thus the equivalent circuit of Fig. 57*B* is obtained. Inductances L_1 and L_2 represent the inductances of the short sections of the center conductors not included in distances d_1 and d_2.

The simplified equivalent circuit of Fig. 57*C* is obtained if the reactance of the open-end resonator line is represented by a series *L*-*C* circuit resonant at the quarter-wave frequency of the line. Capacitance C_a is omitted because, for usual aperture shapes, the susceptance of C_a is

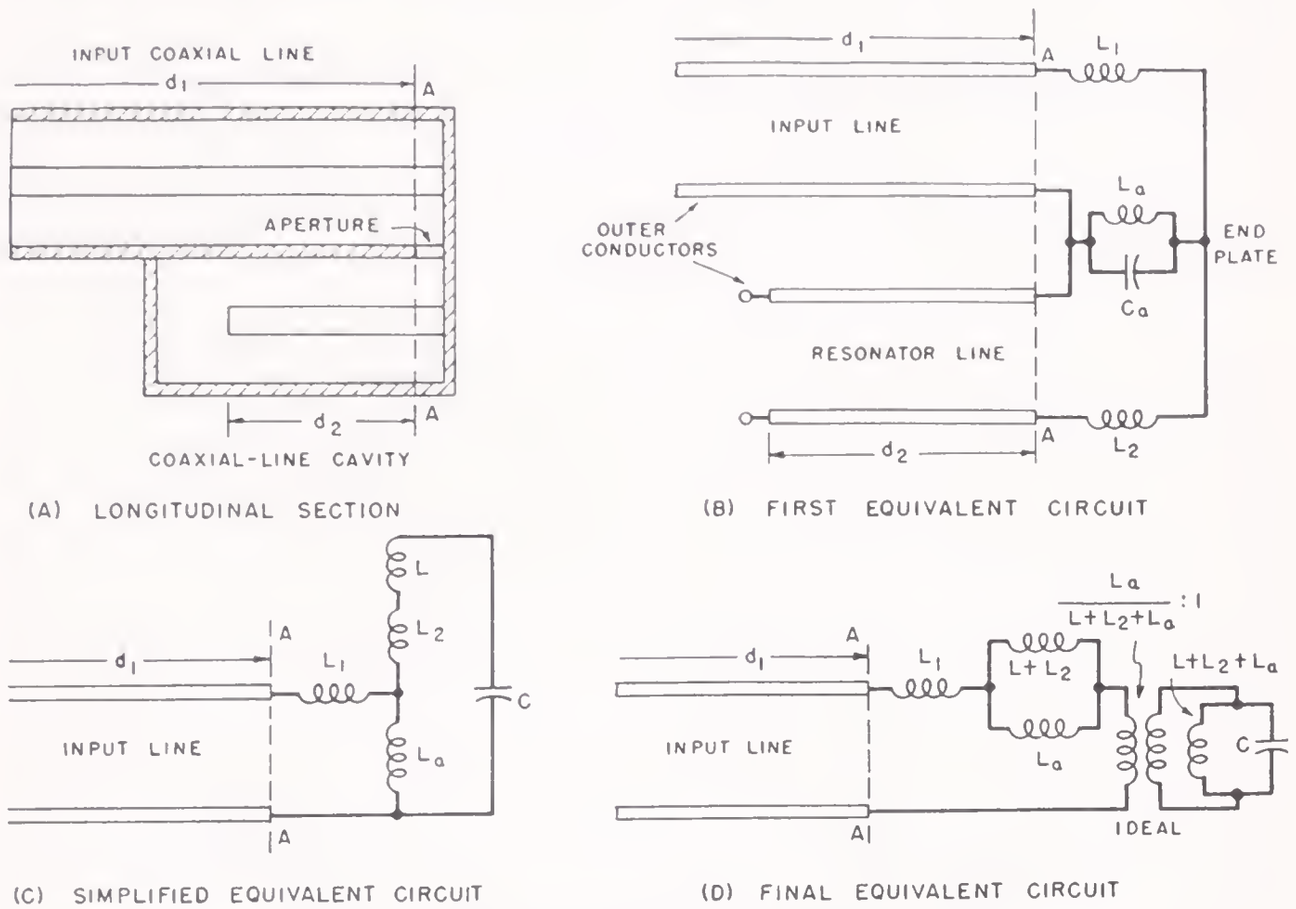


FIG. 57. Aperture coupling in coaxial-line system.

negligible relative to that of L_a. The final equivalent circuit of Fig. 57*D*, obtained from Fig. 57*C*, indicates that the input line is terminated at plane *AA* by a resonator with inductive coupling reactance and that the loaded Q of the system depends upon L_a and therefore upon the dimensions of the aperture.

Sometimes the aperture-coupled coaxial-line system of Fig. 58*A* is employed to place a resonator in series with one conductor of a transmission line. The system of Fig. 58*A* may be obtained by substituting a coaxial line for the input-line end plate in Fig. 57*A*. Therefore an equivalent circuit for Fig. 58*A* is obtained by replacing the short circuit at the end of the input line in Fig. 57*B* with a line section. This equivalent circuit, in simplified form, is shown in Fig. 58*B* and indicates the series relationship of the resonator and the line conductor. In this

circuit L_c is the total coupling inductance, and L' and C' are obtained by referring $L + L_2 + L_a$ and C in Fig. 57*D* to the primary of the ideal transformer.

A waveguide analogue of the coaxial-line system of Fig. 58*A* is sketched in Fig. 59*A*. The rectangular guide is assumed to propagate only $TE_{1,0}$ waves, and the aperture is a rectangle with its long dimension parallel to the magnetic field of the cavity and to the transverse component of

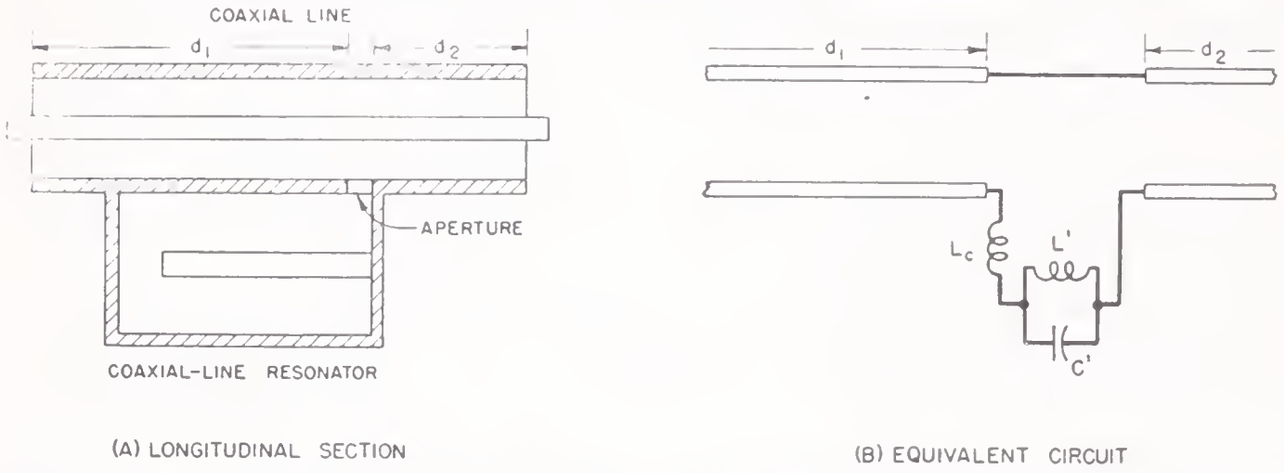


FIG. 58. Resonator connected in series with coaxial lines by means of aperture coupling.

magnetic field in the guide. The transverse-magnetic field is associated with the traveling-wave aspect of the guide; thus the aperture may be considered to interrupt the longitudinal current of the traveling wave and to give rise to a series element in the transmission-line equivalent circuit of Fig. 59*B*. The element values in this circuit have the same

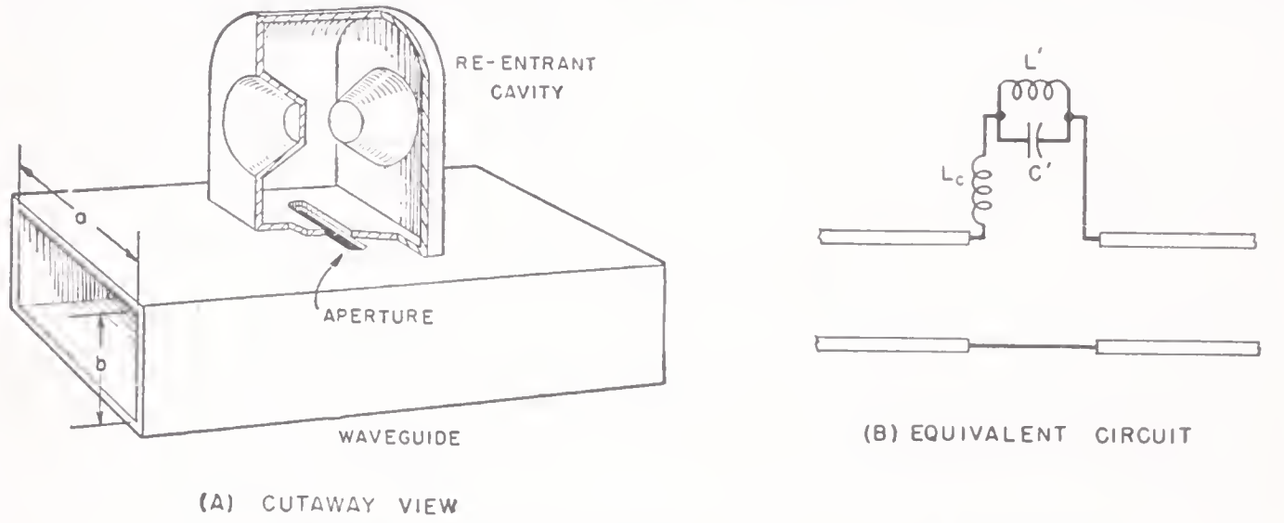


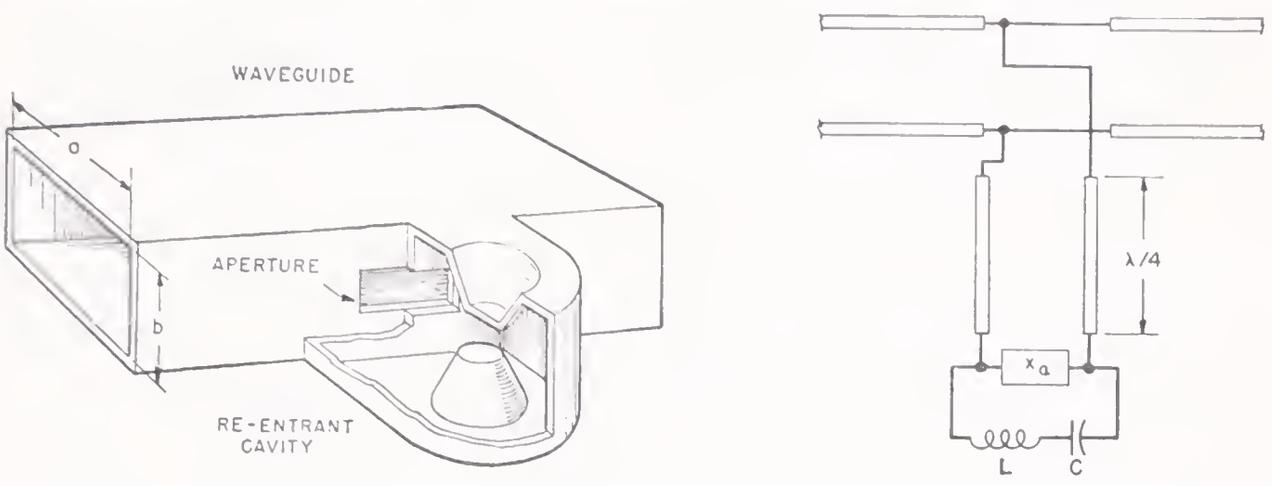
FIG. 59. Resonator coupled by aperture to top wall of a waveguide.

general significance as those in Fig. 58*B*; inductive coupling reactance is indicated because the aperture is assumed to have such a shape that its reactance is inductive.

A more common method of coupling a cavity resonator to a waveguide is illustrated in Fig. 60*A*. A resonant cavity adjacent to one side wall of the guide is coupled to the guide by a rectangular aperture having its long dimension parallel both to the longitudinal component of magnetic

field in the guide and to the magnetic field of the cavity. The interrupted current corresponding to the longitudinal field in the guide is the transverse current in the guide walls associated with the standing-wave aspect of the guide. In its standing-wave aspect, the $TE_{1,0}$ guide wave comprises a half-wave pattern with zero-voltage points at the side walls. Thus the aperture does not affect the guide operation if it introduces zero impedance into the path of the transverse current, but it is equivalent to a short circuit at the guide center if it becomes an open circuit in the transverse-current path.

This effect of the aperture upon the waveguide is represented in the equivalent circuit of Fig. 60*B* by means of a quarter-wave line joining



(A) CUTAWAY VIEW (B) EQUIVALENT CIRCUIT

FIG. 60. Resonator coupled by aperture to side wall of a waveguide.

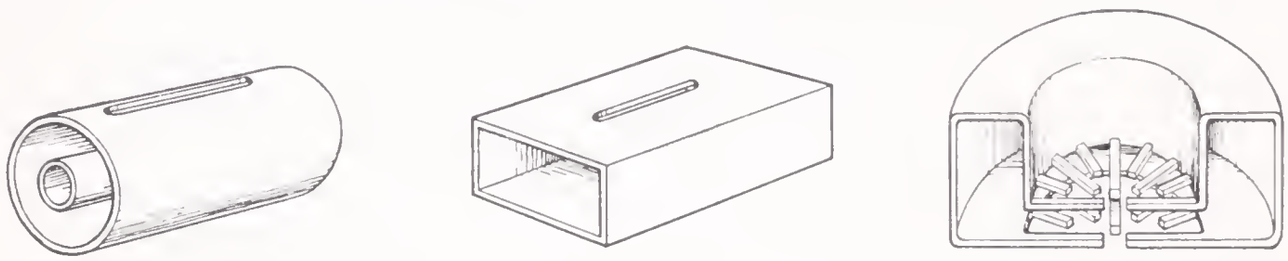
the aperture to the transmission line that represents the guide. The aperture lies in the path of current in the resonator wall as well as in the path of transverse guide-wall current. Thus the aperture reactance X_a in the equivalent circuit connects to the *L*-*C* circuit that represents the resonator as well as to the quarter-wave line.

At the resonant frequency f_0 of L and C, the quarter-wave line is terminated in a short circuit and the aperture and resonator have no effect upon the guide. At most other frequencies, the terminating reactance is essentially equal to X_a, which ordinarily is negligible, and the effect of the resonator on the guide is again small. At some frequency near f_0, however, parallel resonance of X_a with the reactance of the *L*-*C* branch occurs. The quarter-wave line is then terminated by an open circuit, and the aperture and resonator have the effect of a short circuit across the guide.

In the preceding examples, the coupling provided by an aperture has been found to depend upon the magnetic field at the aperture location. An aperture in the region of a cavity wall of weak magnetic field provides very little coupling even though an intense electric field may terminate

upon this portion of the wall. Any coupling that does exist is the result of electric-field lines which penetrate the aperture and terminate upon an external conductor. Such coupling may be thought of as probe coupling made weak by complete withdrawal of the probe (the external conductor) from the cavity.

The smallness of the coupling provided by an aperture that does not interrupt lines of current flow is used to advantage in the devices illustrated in Fig. 61. Figure 61*A* shows the essentials of a slotted section of coaxial line, and a similar representation of a slotted section of wave-



(A) SLOTTED COAXIAL LINE (B) SLOTTED WAVEGUIDE (C) KLYSTRON CAVITY

FIG. 61. Devices containing apertures that provide little coupling.

guide appears in Fig. 61*B*. Because the long, narrow slots are parallel to the lines of current flow, the interrupted current is negligible, and little power escapes from the slots. A small probe may be located in or near the slots and moved along the line or guide to determine the standing-wave pattern in the system (see Art. 18, Chap. XII). The cavity of Fig. 61*C* is typical of the re-entrant cavities used in klystron oscillators (see Chap. XI). The capacitor plates of the resonator are actually grid structures that permit the electron beam to enter and leave the cavity. Little of the oscillation energy can escape from the cavity, because the magnetic-field intensity in the region of the holes is small and the flow of current is parallel to the grid rods shown in the figure.

CHAPTER X

HIGH-FREQUENCY TRIODE OSCILLATORS

Pulsed radar systems contain two generators of high-frequency electrical oscillations. One of these is the *transmitting oscillator*, which provides the powerful high-frequency pulses needed for echo detection. The other is the *local oscillator* of the superheterodyne receiver. This oscillator produces a low-power continuous oscillation of a frequency nearly equal to the transmitting frequency. In uhf radar sets, both the transmitting and the receiving oscillators employ triode tubes. In microwave systems, the transmitting oscillator is usually a magnetron, and the receiving oscillator is a klystron. The triode oscillators of uhf systems are the subject of this chapter; magnetrons and klystrons are discussed in Chap. XI.

1. Ultrahigh-frequency Triodes. If an attempt is made to adjust the frequency of a triode oscillator to higher and higher values, a limiting frequency above which oscillation is impossible is eventually reached. The value of the maximum frequency depends upon the tube and for many standard low-frequency tubes is less than 100 Mcps. One cause of the frequency limit is the existence of minimum values of capacitance and inductance in the circuit. To increase frequency, either capacitance or inductance may be decreased; but the capacitance cannot be made less than the value provided by the interelectrode capacitances of the tube, and the inductance is limited to the lead inductances of the tube plus the inductance of the shortest possible connection between electrode terminals.

Many tubes deliver progressively less power as the frequency is increased and finally fail to oscillate at a frequency below the limit set by the minimum attainable values of capacitance and inductance. One cause of this effect is the low impedance level of the resonant circuit at high frequencies. If the capacitance is reduced to its minimum value, and if further increase of frequency by decrease of inductance is attempted, a decrease of the $\sqrt{L/C}$ parameter of the resonant circuit results. In consequence the load impedance for the tube becomes progressively less and eventually reaches values far too small for efficient operation of the tube. A more fundamental cause of the oscillation-frequency limit is the increased importance of the *transit time* of electrons in the tube at high frequencies. Triode oscillators cannot function

efficiently if the transit time exceeds a small part of the period of oscillation, no matter what adjustments of the resonant-circuit elements are made.

To be useful as a uhf oscillator, a triode must have a short transit time and small interelectrode capacitances. One approach to uhf-triode design is to use the conventional arrangement of electrodes within a glass envelope, to decrease the size and spacing of the electrodes, and to employ short, heavy, well-separated connecting leads. The connecting-lead arrangement contributes to low values of lead inductances, as well as

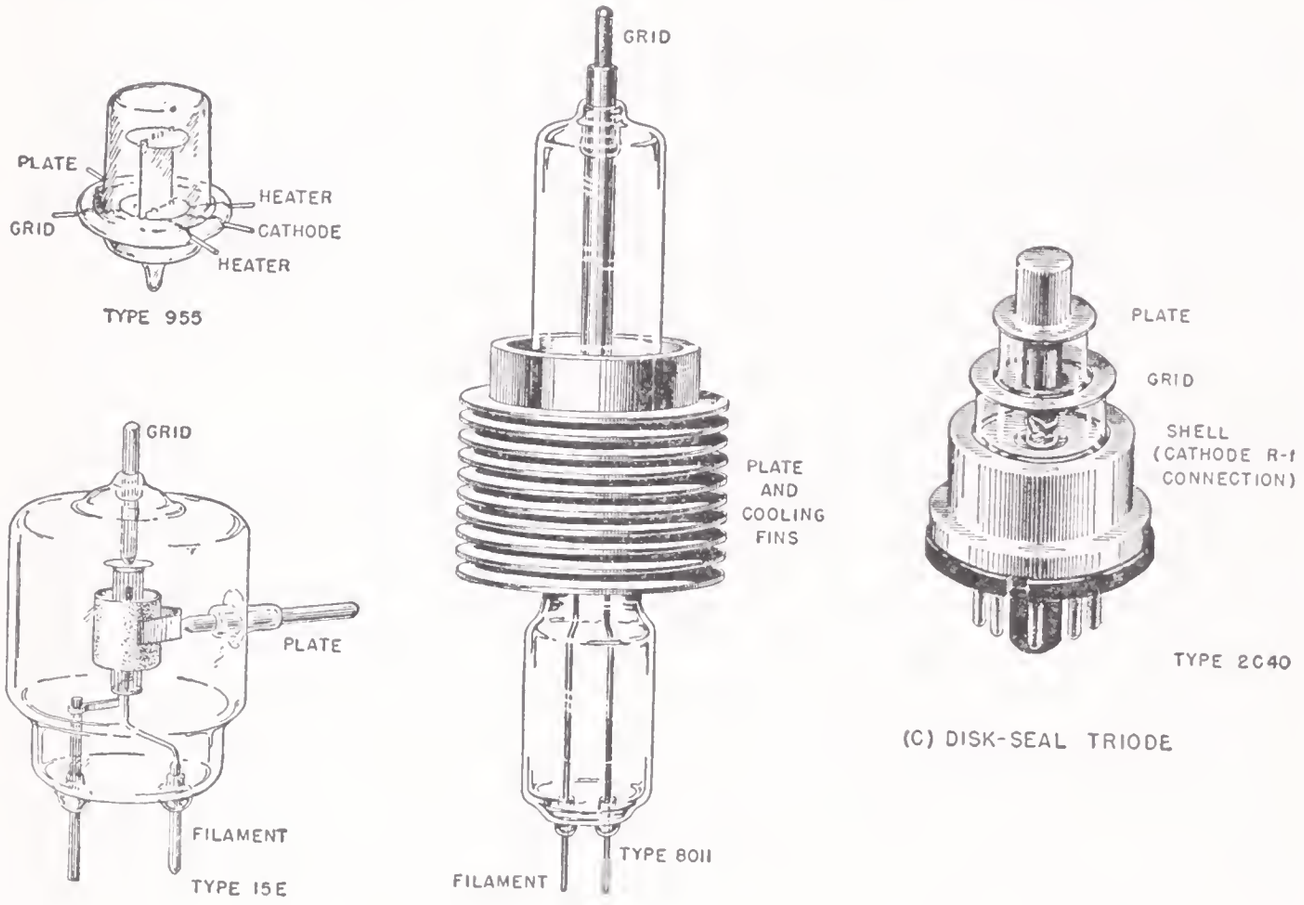


FIG. 1. Typical ultrahigh-frequency tubes (approximately three-fourths of actual size).

minimizing the interlead capacitances. Examples of this design are given in Fig. 1*A*. The Type 15*E* triode is useful in radar transmitters operating at about 500 Mcps and can supply pulses of about 4-kw peak power. The type 955 acorn triode is smaller and can operate at slightly higher frequencies, but it is useful only as a local-oscillator tube in receivers because of its very limited power output.

Because of the transit-time and capacitance requirements, a series of tubes of the general type illustrated by Fig. 1*A* shows a decrease of size and power output with increase of frequency rating. This relation of size and output to frequency is important and occurs in any series of triodes (or of klystrons or magnetrons—see Chap. XI) of a single design.

Any increase of frequency without loss of power output requires a definite design change. Of particular importance for high-frequency performance are plate designs that combine small size with large power-dissipating capacity and cathodes that provide high emission-current densities so that large currents may be possible in small tubes.

The tubes of Fig. 1*B* and *C* illustrate two designs in which improved performance is obtained at the cost of increased difficulty of manufacture. Both of these designs employ large metal-to-glass seals that require special manufacturing techniques. In the tube of Fig. 1*C*, electrode spacings are so small that extreme precision in assembly is required.

The tube of Fig. 1*B* is one of a series of external-anode tubes[1] some of which operate in the region of 600 Meps and supply peak powers as great as 200 kw. The plate is a copper tube forming the central part of the tube envelope and is surrounded by an assembly of cooling fins over which a blast of air is blown to permit dissipation of large powers. The grid is a parallel-rod structure similar to the grid visible in the Type 15E drawing, and the grid and plate are intended to form a continuation of an external grid-anode coaxial line. These tubes are particularly suited to plate-return circuits (Art. 4) in which the plate structure forms part of a wall separating the grid-plate input circuit from the cathode-plate output circuit.

The Type 2C40 disk-seal tube of Fig. 1*C* is called a *lighthouse tube*[2] because of its appearance. This tube utilizes a parallel-plane electrode structure, shown by the cross section of Fig. 2*A*, in which the cathode-grid spacing is only 0.004 in. The 2C40 is useful as a receiver oscillator at frequencies as high as 3,000 Meps and as an r-f amplifier at somewhat lower frequencies. Another lighthouse tube almost identical in appearance, the 2C43, can provide short pulses of about 1-kw peak power at frequencies up to 3,000 Meps. Disk-seal tubes are designed to fit into the end of a coaxial-line tuner (Art. 7) and are particularly adapted to a grid-return circuit (Art. 6) in which the grid and grid ring of the tube form part of a wall that separates the cathode-grid input circuit from the plate-grid output circuit.

Disk-seal tubes are made in considerable variety;[3] two possibilities in addition to the lighthouse tubes are suggested by Figs. 2*B* and *C*. In

[1] J. BELL, M. R. GAVIN, E. G. JAMES, and G. W. WARREN, "Triodes for Very Short Waves—Oscillators," *J.I.E.E. Part IIIA*, **93** (March–May, 1946), 833–846.

[2] E. D. MACARTHUR, "Disk-seal Tubes," *Electronics*, **18** (February, 1945), 98–102.

[3] H. W. JAMISON and J. R. WHINNERY, "Power Amplifiers with Disk-seal Tubes," *Proc. I.R.E.*, **34** (July, 1946), 483–489; J. J. GLAUBER, "Radar Vacuum-tube Developments," *Elec. Comm.*, **23** (September, 1946), 306–319; R. R. LAW, D. G. BURNSIDE, R. P. STONE, and W. B. WHALLEY, "Development of Pulse Triodes and Circuit to Give One Megawatt at 600 Megacycles," *RCA Rev.*, **7** (June, 1946), 253–264.

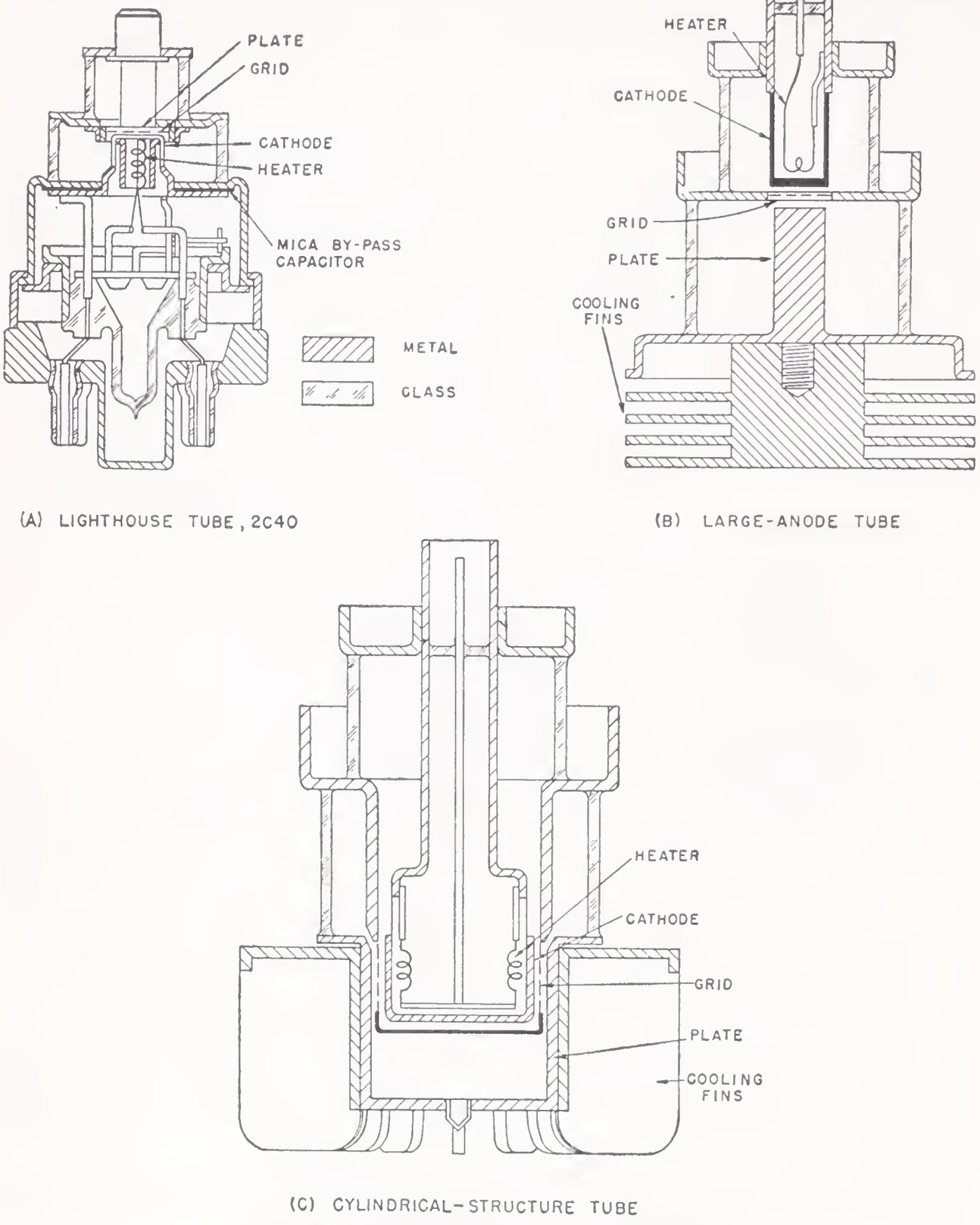


FIG. 2. Cross sections of disk-seal tubes.

Fig. 2*B* the plate is located at the large end of the tube and equipped with a cooling structure. (Cooling of the lighthouse-tube anode is difficult, because the anode is located in the center of the coaxial tuner.) A tube structure very similar to Fig. 2*B*, but adapted to higher power, makes use of a cylindrical electrode geometry, as shown in Fig. 2*C*. With such tubes, peak pulse powers of the order of 500 kw are possible at 600 Mcps.

A great variety of different uhf triodes is employed in radar systems, but nearly all are variations of the three types of construction exemplified by the tubes of Fig. 1. Occasionally two or more triode assemblies, together with associated tuning elements, are built into a single evacuated envelope.[1]

The way in which the operation of uhf oscillators is impaired by large transit time is suggested in a qualitative manner by Arts. 9 and 11 of Chap. VI. However, the discussions of these articles do not apply in detail to oscillators, because in Chap. VI only the small-signal operation of tubes was considered. In oscillators, high-efficiency Class C operation is employed, and plate current flows in pulses of less than a half-cycle duration. In a 500-Mcps oscillator, for example, the oscillation period is 0.002 μs and the duration of each pulse of plate current may be only 0.0005 μs. The operation of the tube depends largely upon the relation of the transit time to this very brief time of current flow.[2]

Consider first the movement of electrons from cathode to grid. Except during a short fraction of the cycle, the grid voltage is below cutoff and the interelectrode space is essentially free of electrons. When the grid voltage first rises above cutoff, electrons start from the cathode; and if the grid voltage remains above cutoff for a long enough time, the electrons pass between the grid wires and eventually reach the plate. However, those electrons which leave the cathode immediately before the grid voltage returns to a large negative value may be stopped and returned to the cathode. This return of emitted electrons is called *back bombardment* and may cause very appreciable heating of the cathode, because the electrons strike the cathode with considerable velocity.

The heating effect of back bombardment is not in itself harmful, provided the heater current is reduced to maintain the correct cathode temperature, but with it are associated two undesirable effects. One is the waste of cathode emission represented by the returned electrons. The current to the plate and the power developed are less than would be possible if the entire cathode emission could be utilized. The other effect is the waste of r-f input power associated with back bombardment. The energy of the returned electrons is derived from the grid-cathode electric field, and thus power supplied by the source of the grid-cathode r-f signal is wasted in heating the cathode.

The movement of electrons through the grid-plate space causes a cur-

[1] H. A. ZAHL, J. E. GORHAM, and G. F. ROUSE, "A Vacuum-contained Push-pull Triode Oscillator," *Proc. I.R.E.*, **34** (February, 1946), 66W–69W; C. E. FAY and J. E. WOLFE, "A Tunable Vacuum-contained Triode Oscillator for Pulse Service," *Proc. I.R.E.*, **36** (January, 1949), 234–239.

[2] CHAO-CHEN WANG, "Large-signal High-frequency Electronics of Thermionic Vacuum Tubes," *Proc. I.R.E.*, **29** (April, 1941), 200–214.

rent in the plate circuit during the entire period of the grid-plate flight rather than a sudden pulse at the instant of electrons striking the plate. The reason is that electrons approaching the plate cause an increasing positive charge to be induced on the plate surface and the arrival of the electrons at the plate neutralizes this charge (see Art. 11, Chap. VI). In consequence, the duration of the current pulse in the plate circuit is greater than the period during which electrons strike the plate by the amount of the grid-plate transit time. For high-efficiency operation, plate current should flow only while the plate voltage is at its minimum value. In this way, very little of the power from the plate supply is lost as anode dissipation; most of the supply power is converted to high-frequency output power. Transit-time lengthening of the plate-current pulse results in a higher plate voltage during a portion of the period of current flow and therefore causes a loss of efficiency.

These transit-time considerations require that both the cathode-grid and grid-plate spacings be small in the tubes of Figs. 1 and 2. The transit times depend not only upon the spacings, however, but also upon the velocities of the electrons. In the cathode-grid region, electrons start with nearly zero speed and are accelerated by the electric field. For short cathode-grid transit time, therefore, the accelerating field must be strong. In space-charge-limited operation, high field intensities can occur only if the density of the electron current is great. Thus a cathode capable of supplying a high current density is as essential as is close electrode spacing to the attainment of short cathode-grid transit time. For this reason oxide-coated cathodes are preferred in uhf tubes; the tubes of Fig. 2 all employ oxide-coated cathodes.

Short grid-plate transit times require, in addition to close electrode spacings, a high value of plate voltage during the time of plate-current flow. Efficient operation requires that this plate-voltage value be a small fraction of the supply voltage. Therefore, very high supply voltages are desirable if grid-plate transit time is to be small.

The requirements of high current and high voltage in a tube of small size produce very difficult anode-cooling problems in high-frequency power triodes. The pulsed operation of radar transmitters is helpful in this respect because of the long cooling periods available between pulses. Cooling problems occur not only for the anode but also for the grid. The grid is heated by electron bombardment (when it is positive at the peak of the grid-voltage wave) and by radiation from the cathode. The expansion of the grid wires at high temperature must be allowed for in the tube design, and if the temperature becomes too high, the grid may become an electron emitter. Grid emission depends greatly upon the surface conditions of the grid wires and is an especially important con-

sideration in tubes with oxide-coated cathodes—some of the cathode material is evaporated onto the grid during operation of the tube.

2. Sustained Oscillations. A great variety of vacuum-tube oscillators—low-frequency and high-frequency triode, tetrode, and pentode circuits and klystrons and magnetrons, for example—depends upon the same principle of operation. Each of these oscillators contains a *resonator*, which may be an *L-C* circuit, a transmission-line section, or a cavity resonator (depending upon the oscillation frequency—see Chap. IX). Associated with the resonator is the *electron beam* of a vacuum tube. If it were not for the electron beam, oscillations started in the resonator would soon die away because of losses in the resonator and because energy in the resonator is delivered to a load. The motion of electrons in the beam is controlled in such a way that the power removed from the resonator is continually replenished by the electrons, and oscillations continue at constant amplitude.

In low-frequency oscillators, the resonator elements and the vacuum tube containing the electron beam form completely separate parts of the circuit. In klystrons and magnetrons, the two components are usually built into the same vacuum system and are closely interrelated. Ultra-high-frequency triode oscillators may be considered an intermediate case. The major portion of the resonator is ordinarily external to the tube; yet, particularly with disk-seal triodes, elements of the tube are important to the resonator as well as to the electron beam.

In conjunction with the frequency-determining resonator and the energy-supplying electron beam, an oscillator must contain a *control mechanism* to regulate the movement of electrons in synchronism with the resonator oscillations in such a way that the electrons may deliver energy to the resonator. In a triode oscillator, the control results from a feedback connection from the resonator to the grid circuit of the triode. The feedback voltage controls the flow of electrons from the cathode, and these electrons provide a plate current having an a-c component in the proper phase to deliver power to a resonator connected as a plate-circuit load.

An oscillator that illustrates these principles in a particularly simple fashion is the tuned-plate circuit of Fig. 3*A*. This circuit is idealized in that losses in the *L-C* resonator and tube capacitances and inductances are ignored, the transformer is considered ideal, and transit time is neglected. These idealizations can very nearly be realized in circuits of moderately low frequency. In order to simplify the discussion, no load is shown connected to the oscillator of Fig. 3*A*.

If an oscillation is started in the resonator, the plate voltage e_b, being less than E_{bb} by the amount of the resonator voltage e_L, has the waveform

indicated in Fig. 3*B*. The sinusoidal voltage e_L, multiplied by the transformer ratio a, is transferred to the grid circuit; and because of the grid-leak bias built up on capacitor C_c the grid voltage e_c has the waveform shown. Because of the large magnitude of the bias voltage, plate current is cutoff during a large portion of the cycle and flows only when the grid voltage exceeds the cutoff value e_{co}. In Fig. 3*B*, the variable cutoff voltage e_{co} (rather than a fixed voltage E_{co}) is shown because the variation of e_b causes a corresponding variation of the cutoff voltage. The pulses of plate current that result are shown in the figure. The waveforms

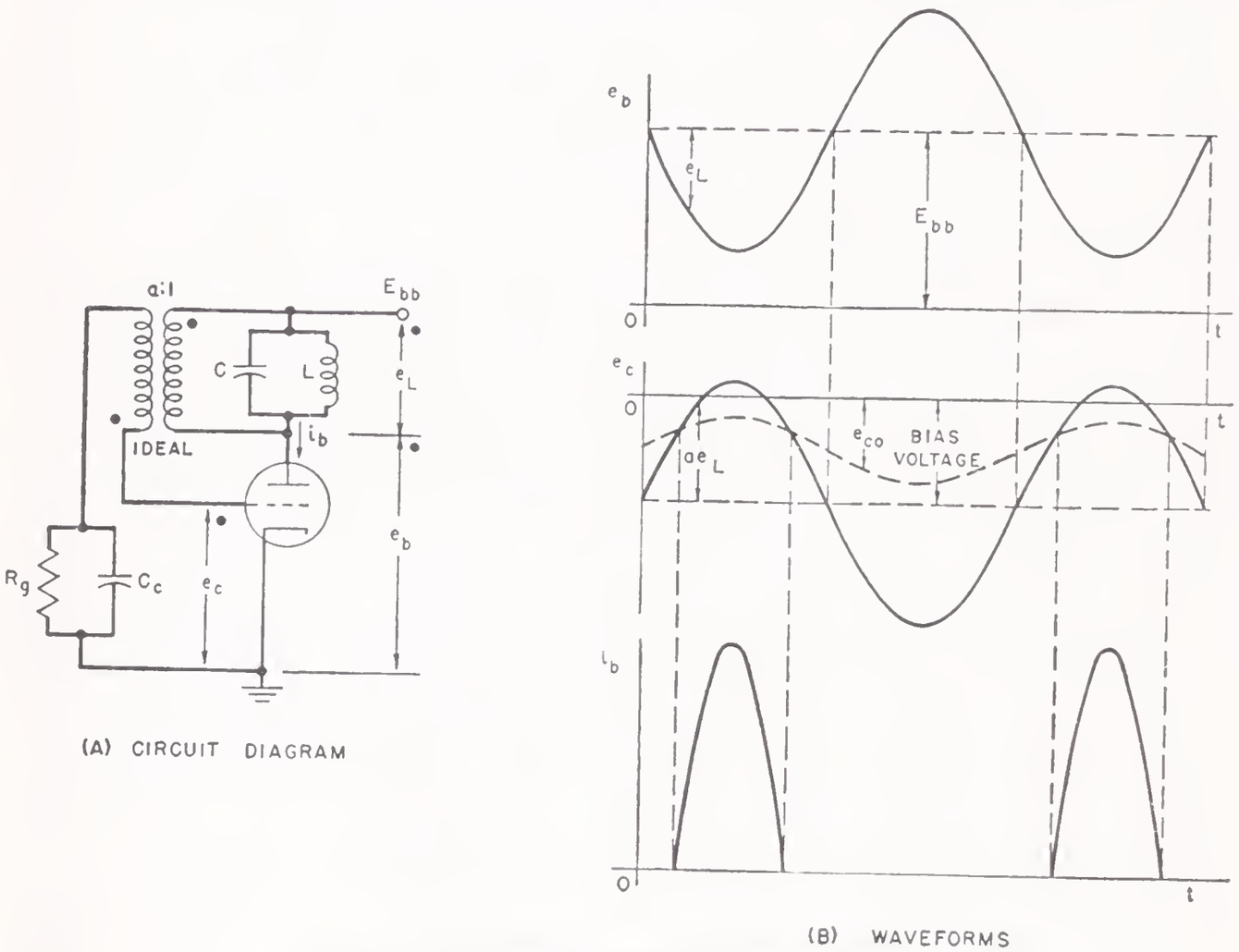


FIG. 3. Low-frequency tuned-plate oscillator.

drawn correspond to Class C operation, the period of current flow being less than half of the cycle.

The phase of the resonator voltage at the time of plate-current flow is all-important to the supply of power to the resonator. The power delivered by the source of E_{bb} at any instant is E_{bb} times the plate current i_b at that instant. For the conditions of Fig. 3*B*, this power divides between the resonator and the tube, the portion $e_L i_b$ going to the resonator and the portion $e_b i_b$ being wasted in the tube as plate dissipation. Had the pulse of current occurred in the half cycle during which e_b exceeds E_{bb}, both the E_{bb} source and the resonator would have supplied power to the tube and the plate dissipation would have been very great. The tube

would then have damped the resonator oscillations instead of sustaining them. For high-efficiency operation, e_b should reach a very low value during one portion of the cycle, and the feedback and bias should be adjusted so that i_b flows only while e_b is near its minimum value. Under these conditions, the plate dissipation is small and nearly all the power from the plate supply is delivered to the resonator.

A very useful point of view, particularly in connection with high-frequency tubes, is to consider the electric field in the tube that results from the plate voltage to be divided into two components, one the result of the supply-voltage component E_{bb} of the plate voltage and the other the result of the resonator-voltage component $-e_L$. The field component resulting from E_{bb} is independent of time and accelerates the electrons toward the plate. If this component alone were present, the electrons would strike the plate with high velocity, and all the power from the E_{bb} source would be converted to heat at the plate. The field component resulting from $-e_L$ is alternating and may accelerate or decelerate the electrons, depending upon the phase of e_L at the time of plate-current flow. For the conditions of Fig. 3*B*, the alternating-field component decelerates the electrons, and the energy lost by the electrons is received by the oscillations in the resonator. Thus the tube may be visualized as a means for first changing the d-c energy of the E_{bb} source to kinetic energy of electrons in the tube and then changing part of the kinetic energy to a-c energy in the resonator. Efficient operation is obtained if the deceleration produced by the a-c field nearly equals the acceleration resulting from the d-c field, so that but little kinetic energy is lost as heat when the electrons strike the plate.

Some of the adjustments that are of basic importance in the operation of all vacuum-tube oscillators may be explained with reference to the circuit of Fig. 3. The first adjustment is of the natural frequency of the resonator and may be an adjustment of L or C in Fig. 3*A*. This adjustment controls the frequency of oscillation. The second and third adjustments control the feedback voltage or the *excitation* of the oscillator. Two excitation adjustments are needed in order that both the phase and the amplitude of the plate-current pulse may be controlled.

In Fig. 3 only two phase positions of the current pulse, corresponding to the two directions of connecting to the transformer, are possible, but one of these provides the 180-deg relation between grid and plate voltage desired in low-frequency oscillators. In high-frequency oscillators, transit time may make a different phase relation between the voltages desirable, and a smooth control of phase is therefore preferred. The excitation-amplitude adjustment in Fig. 3*A* is the transformer ratio a. Oscillations are possible only if a exceeds a certain minimum value, and the strength of oscillations and efficiency of operation increase as a is

made somewhat larger than this minimum value. Too large a value for a, however, increases the losses in the grid circuit and results in inefficient operation. Adjustment of R_g influences the bias voltage and should be considered in conjunction with adjustment of a; R_g and a together determine the amplitude and duration of the current pulse.

The circuit of Fig. 3*A* does not include a load for the oscillator and therefore contains no means for adjusting the loading of the circuit. To be useful, an oscillator must deliver power to a load, and the loading control is a fourth basic adjustment. Methods of applying and adjusting loads for oscillators are considered in Arts. 8 to 10.

3. Ultra-audion Oscillators. One of the oldest and simplest of the high-frequency oscillator circuits comprises essentially an inductance connected between grid and plate of a triode of conventional design. The grid-plate inductance and the tube capacitances form the frequency-determining resonator, and the tube capacitances supply the excitation. The oscillator of Fig. 4, in which L is the inductance of a few turns of heavy copper wire, is useful in low-power applications (such as radar-receiver local oscillators) at frequencies up to, roughly, 200 Mcps. The similar circuit of Fig. 6, in which inductances are replaced by transmission-line sections, is useful at somewhat higher frequencies.

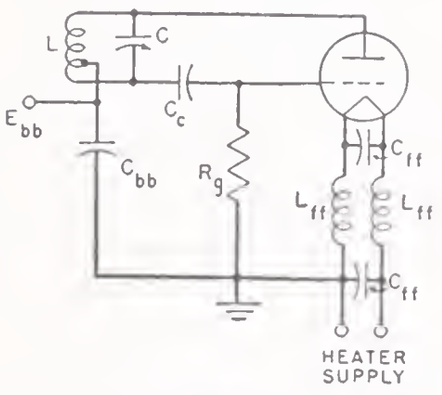


FIG. 4. Ultra-audion oscillator employing L-C resonator.

In Fig. 4, capacitor C is a very small adjustable capacitor and is included in the circuit merely as a means of controlling the frequency of oscillation. Elements C_c and R_g permit the development of grid-leak bias. Capacitors C_{bb} and C_{ff} are low-impedance by-pass capacitors, and inductors L_{ff} are r-f chokes having negligible admittance at the oscillation frequency.

An r-f equivalent circuit in which stray circuit elements are shown explicitly is drawn in Fig. 5*A*. The oscillator components are assumed to be mounted within a metal shielding enclosure, and the shield is considered to be at ground potential. The capacitances between the two ends of the resonator coil and the shielding metal are denoted by C_1 and C_2. The tube capacitances are indicated as C_{gp}, C_{gk}, and C_{pk}, and the tube lead inductances are L_g, L_p, and L_k.

In all high-frequency oscillators, means must be provided for preventing r-f currents from entering the power-supply circuits. Diversion of r-f power to the supply circuits would decrease the useful output of the oscillator; also r-f currents in the power supply might interfere with the operation of other circuits connected to the same supply. In Fig. 5*A*, the filter comprising capacitors C_{ff} and inductors L_{ff} provides heater-

supply isolation and at the same time permits a large r-f voltage to exist between cathode and ground. Plate-supply isolation is obtained by means of C_{bb} and the location of the plate-supply tap on L. If the circuit were to oscillate with C_{bb} open, a point such as A on L would have the same potential as point B (see Fig. 5A), which is on the shielding structure. Thus no r-f current enters the power supply if the tap is connected to point A. The low reactance of C_{bb} between points A and B gives additional assurance of plate-supply isolation.

The circuit of Fig. 5A may be reduced to the simpler circuit of Fig. 5B if only r-f components of current and voltage are considered. At the oscillation frequency, inductances L_{ff} and resistor R_g may to a good approximation be considered open circuits and C_c and C_{bb} are effectively

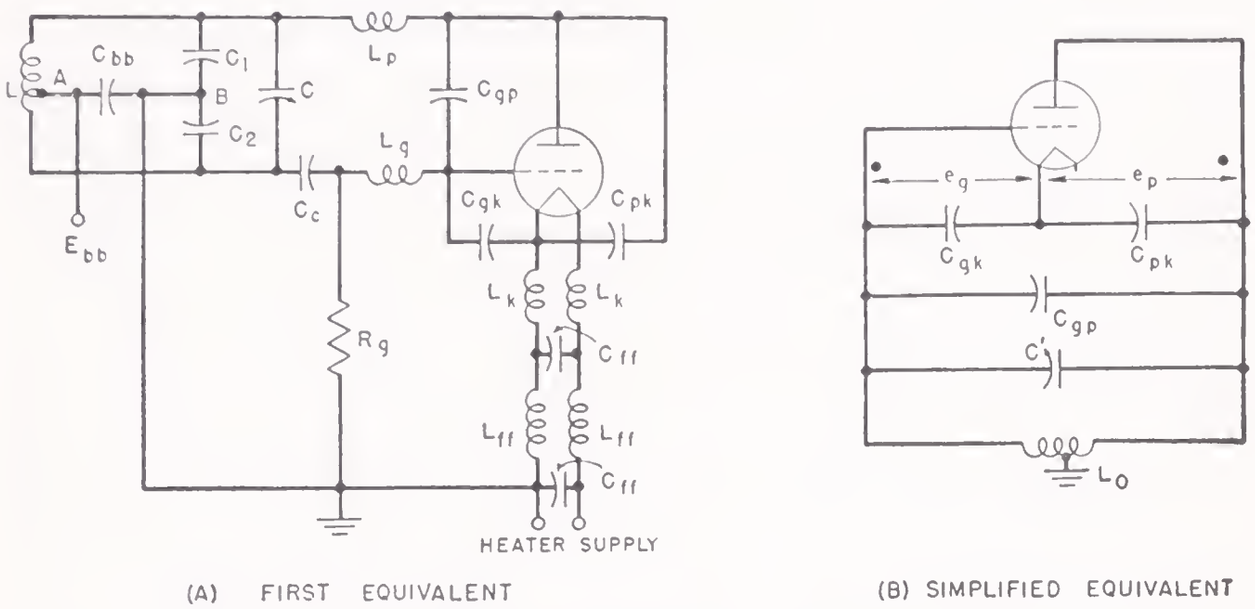


FIG. 5. R-f equivalent circuits for the oscillator of Fig. 4.

short circuits. Suppose for a moment that capacitors C, C_1, and C_2 are not present, and designate the series combination of L, L_p, and L_g as L_0. The circuit of Fig. 5B, but without C', is then obtained. Capacitors C_1, C_2, and C are very small and can with little error be transferred by the resonant-transformer relation (Art. 11, Chap. IX) from their actual position across L to a position across L_0; thus the variable capacitor C' in Fig. 5B is obtained. The final equivalent circuit indicates that the ultra-audion circuit is a Colpitts oscillator in which the tuned-circuit capacitances are reduced essentially to the unavoidable tube capacitances.

The frequency-determining resonator in Fig. 5B is the combination of L_0 with the capacitances C', C_{gp}, C_{gk}, and C_{pk}. The frequency of oscillation is therefore the frequency at which inductance L_0 resonates with the total capacitance

$$C_0 = C' + C_{gp} + \frac{C_{gk}C_{pk}}{C_{gk} + C_{pk}}$$

and the frequency is controlled by adjustment of capacitor C, which in turn controls the value of C'.

The resonator of Fig. 5B has two points of entry—the terminal pair across which the a-c component of plate voltage e_p is applied and the terminal pair at which the a-c component of grid voltage e_g is developed. Thus the resonator, considered as a resonant transformer, provides a grid-voltage wave 180 deg out of phase with the plate-voltage wave and a voltage ratio (determined by the capacitance voltage divider) given by

$$\frac{E_g}{E_p} = \frac{C_{pk}}{C_{gk}}$$

This ratio has an effect on oscillator operation analogous to the transformer ratio a in the circuit of Fig. 3A.

A fundamental disadvantage of the circuit of Fig. 4 is the lack of a means of controlling either the phase or the magnitude of the feedback voltage. The phase angle between the e_g and e_p waves would be exactly 180 deg if the grid current of the tube in Fig. 5B were zero. Actually, however, grid current exists because the grid voltage becomes positive for brief periods in order to maintain the grid-bias voltage and because of transit-time effects. The grid current must be supplied through the capacitive coupling reactances of the resonator (Art. 7, Chap. IX), and therefore the phase angle is not exactly 180 deg. Because of transit time, the desired phase angle also is different from 180 deg, but no adjustment is available to make the actual phase angle equal the desired phase angle.

An alternate ultra-audion circuit represented in semipictorial fashion in Fig. 6 differs from the circuit of Fig. 4 in that the resonator inductance is replaced by a section of transmission line less than a quarter wave in length and the heater-circuit filter is replaced by a quarter-wave coaxial line. The capacitance C between the adjustable disks on the grid-plate line serves to control the frequency of the oscillator. Radio-frequency equivalent circuits essentially the same as those in Fig. 6 may be drawn.

The grid-plate line is perhaps best thought of as a resonant cavity comprising the parallel rods inside the shielding enclosure. If resistor R_{bb} were short-circuited, this cavity would have two natural modes of nearly equal frequency. In the first mode, which is the desired mode, the voltages of the two line conductors are balanced with respect to the shield, one voltage being positive when the other is negative. In the second mode, the two conductors are at the same potential and together may be considered the center conductor of a line for which the shield is the outer conductor. Resistor R_{bb} is chosen to approximate the characteristic resistance of the "coaxial" line of the undesired mode in order to ensure that oscillation in this mode cannot be sustained. The voltage

at the shorting bar of the grid-plate line in Fig. 6 resulting from an oscillation in the desired mode is zero because of the balanced construction of the line. Consequently, isolation of the plate supply circuit is easily obtained.

Precautions against wrong modes of operation or *parasitic oscillations* are necessary in nearly all uhf oscillators, especially those employing more than one tube (see Art. 5), and resistors such as R_{bb} are often called *parasitic suppressors*. Figure 6 illustrates the use of damping to suppress a wrong mode. Another method is to arrange the circuit so that oscillation-sustaining feedback is not provided for the undesired mode.

The concentric cylinders of the heater-supply assembly form a quarter-wave coaxial line that permits a low-frequency connection to the cathode

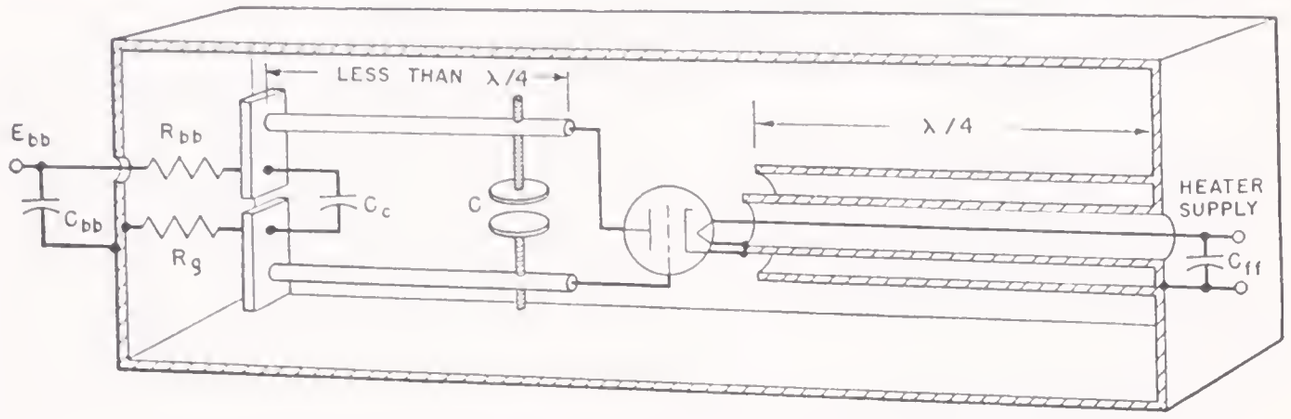


FIG. 6. Ultra-audion oscillator employing transmission-line resonator.

and yet provides an r-f open circuit between cathode and ground. The completion of the heater circuit through a lead inside the center conductor of this line is an example of a very common and useful method of supplying heater power to tubes in high-frequency circuits. The coaxial line formed by the heater lead and the inner cylinder of the cathode line is effectively short-circuited at both ends—at one end by the tube heater and at the other end by capacitor C_{ff}. Thus the line, which is approximately a quarter wave in length, is nonresonant, and very little of the r-f power can be transmitted through it to the heater supply.

4. Plate-return Oscillators. An oscillator circuit in which tuning elements are connected from grid to plate and from cathode to plate is especially suitable with tubes of the external-anode design of Fig. 1*B*. This circuit is variously known as a *plate-return, plate-separation, grounded-plate,* and *tuned-grid, tuned-cathode oscillator* and has been widely used in radar transmitters not only with external-anode tubes but also with triodes of conventional design.

The single-tube plate-return oscillator of Fig. 7 is considered in this article because it illustrates the principles of the circuit in simple fashion. (Because of power requirements, the circuits generally employed in radar transmitters are the multitube circuits of the following article.) In

Fig. 7 an external-anode triode is fitted into a coaxial-line tuner, and direct voltage is applied to the plate-cathode circuit between the E_{bb} terminal and the heater-transformer center tap. The spacers on either side of the anode cooling structure provide high-voltage d-c insulation and at the same time are the dielectrics of two capacitors having negligible reactance at the oscillation frequency. The d-c grid circuit includes the outer conductors of the coaxial lines and resistor R_g. Flow of grid current in R_g causes a bias voltage to be stored on capacitor C_c. The heater-supply circuit is similar to the one in Fig. 6.

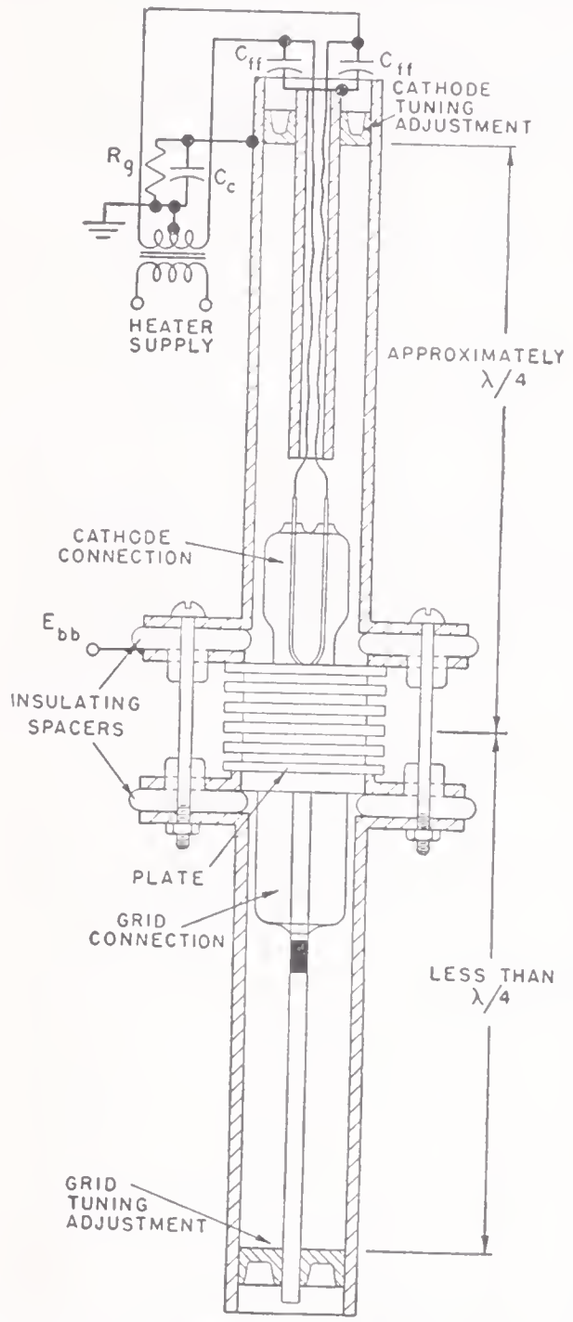


FIG. 7. Single-tube plate-return oscillator.

The tuner of Fig. 7 comprises two coaxial-line resonators separated by the plate structure. The only coupling between these resonators results from the electron beam and the grid-cathode capacitance of the tube. Because of this construction, the plate-lead inductance (which would give rise to a reactance common to the grid and cathode circuits) is zero, and grid and cathode lead inductances may be ignored if the connecting leads are considered part of the coaxial-line center conductors. An r-f equivalent circuit for the oscillator may therefore be drawn as in Fig. 8*A* for the frequency range near the quarter-wave resonances of the two coaxial lines. In this figure, C_k and L_k represent the cathode line (see Art. 8, Chap. IX) and L_g and C_g represent the grid line.

The resonator of Fig. 8*A* has two modes of resonance. These modes may be seen from the diagrams of Fig. 8*B*, in which X_k represents the net reactance of L_k, C_k, and C_{pk} in parallel and X_g is the net reactance of L_g, C_g, and C_{gp} in parallel. The resonant frequency of the cathode line loaded by C_{pk} (the pole of the reactance function X_k) is designated by f_k, and the resonant frequency of the grid line and C_{gp} (the pole of X_g) is denoted by f_g. The natural frequency of the complete resonator is called f_0. For the low-frequency mode of resonance, f_0 is less than both

f_k and f_g, so that both X_k and X_g are inductive, and the series combination of X_k and X_g resonates with C_{gk}, as indicated by the top diagram of Fig. 8*B*. For the high-frequency mode, f_0 lies between f_k and f_g, so that either X_k or X_g is inductive and the other is capacitive. The two conditions possible, depending upon whether the cathode or the grid line is tuned to the higher frequency, are shown by the two lower diagrams of Fig. 8*B*.

Examination of Fig. 8*B* shows that the phase relation between grid and plate voltages required for sustaining oscillations is obtained only if resonance occurs in the high-frequency mode and f_k is less than f_g. The

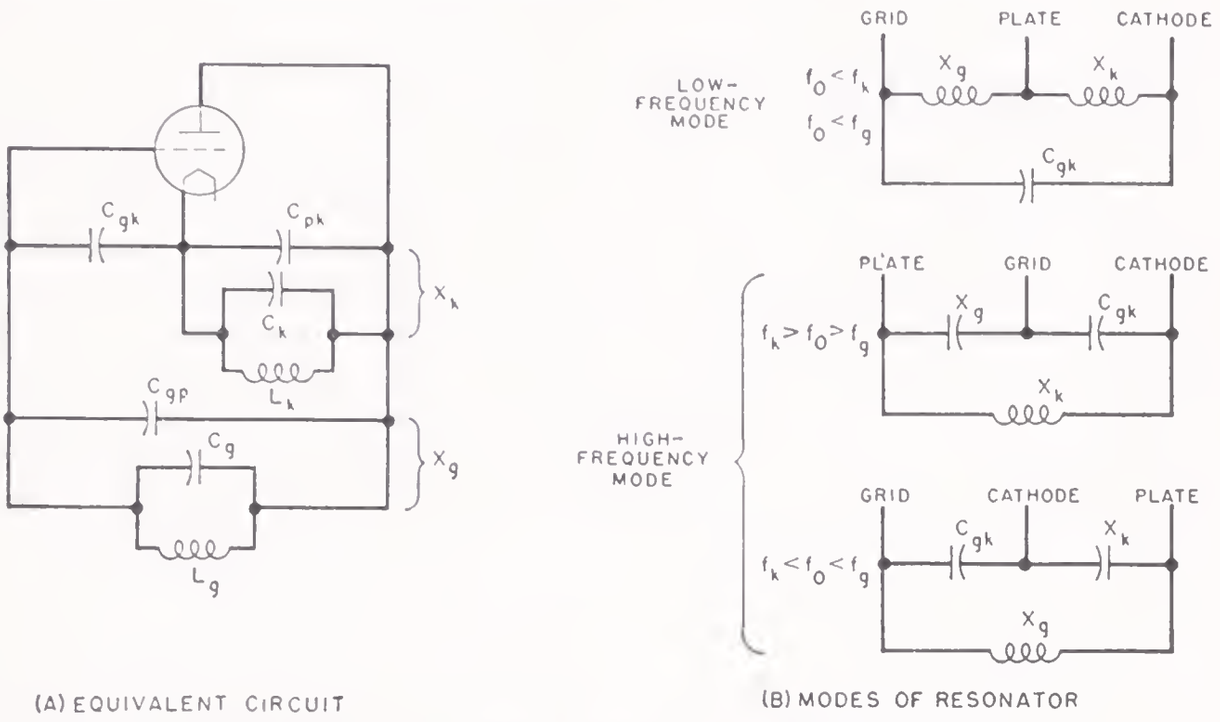


FIG. 8. R-f equivalent circuit for oscillator of Fig. 7.

lumped-parameter equivalent circuit of Fig. 8*A* can therefore oscillate in only this one way. Other modes of sustained oscillation, corresponding to multiple quarter-wave resonances of the transmission lines, are possible in the actual circuit of Fig. 7. Ordinarily, only oscillations of the lowest possible frequency are obtained, however, because transit time becomes more important in the higher frequency modes, and the efficiency of the tube is reduced.

Because X_k is capacitive and X_g inductive at the frequency of oscillation, the equivalent circuit of the plate-return oscillator has the basic Colpitts form. The circuit differs from that of the ultra-audion oscillator only in that the amplitude of the feedback voltage can be adjusted through control of X_k by the cathode-line tuning. If the cathode line is appreciably shorter than the grid line, X_k is inductive and oscillation is impossible. Lengthening the cathode line makes X_k capacitive, but oscillations may still not occur if the magnitude of X_k is too large relative to $1/\omega C_{gk}$, so that the capacitance voltage divider supplies very little

voltage to the grid. Further increase of cathode-line length decreases the magnitude of X_k, and oscillations become possible. Too long a cathode line results in excessive grid voltage and inefficient operation because of the large grid-circuit losses.

The grid-line shorting plunger is conveniently thought of as the frequency control for this oscillator, and the cathode-line plunger as the feedback control. The two controls, however, are not independent, and a succession of adjustments is usually needed to obtain the best operation at a specified frequency. No provision for feedback-phase adjustment is available. However, if operation too near the high-frequency limit is not attempted, and if the r-f grid voltage is not excessive, both the desired and the actual phase angles will approximate 180 deg and the loss in performance will be small.

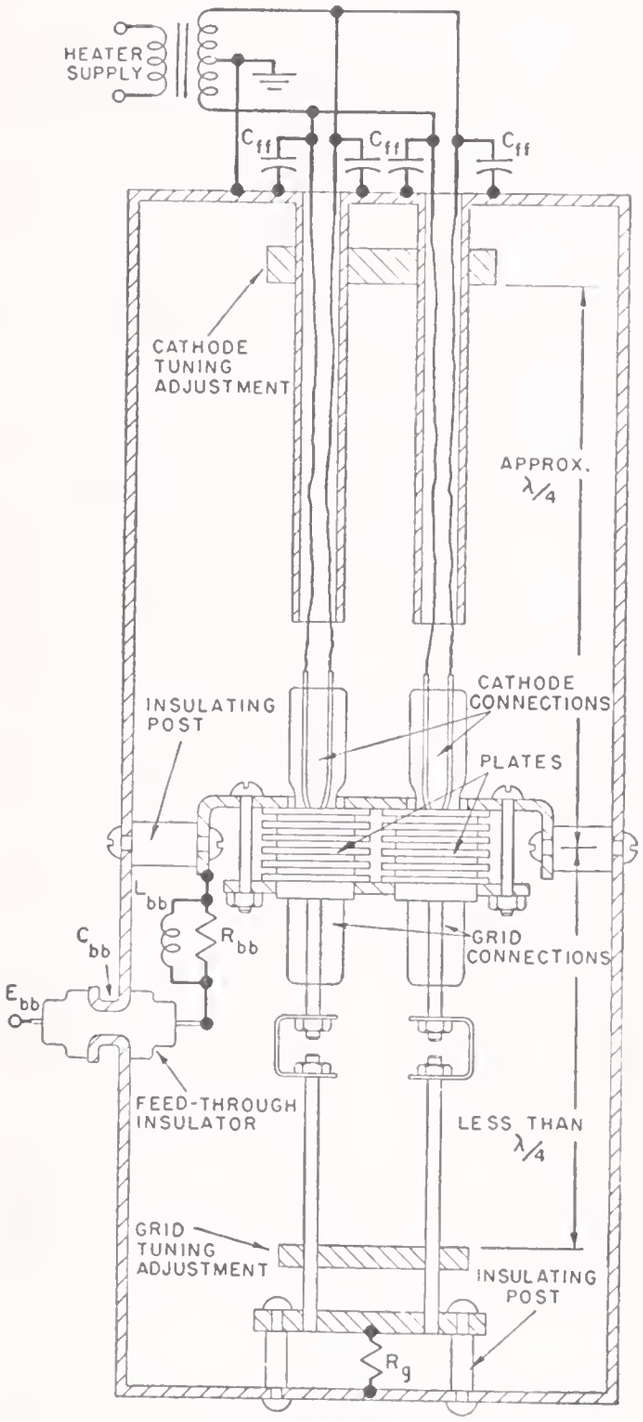


FIG. 9. Push-pull plate-return oscillator.

5. Multitube Oscillators. Because of the tendency of triodes to decrease in size and power capacity with increase of frequency rating, multitube circuits are widely used in uhf power oscillators. *Push-pull circuits* make available approximately twice the power output of a single tube, and higher powers can be obtained from *ring oscillators*, which may comprise any even number of tubes. Unfortunately the possibilities for parasitic oscillations and the difficulty of adjusting the circuit for optimum operation increase as the number of tubes increases. Push-pull circuits are thoroughly practical, but circuits containing more than two tubes are so difficult to adjust that the full power capacity of the tubes is seldom realized.

A push-pull version of the plate-return oscillator of Art. 4 is sketched in cross section in Fig. 9. This circuit may be derived from that of Fig. 7 if the center conductors of the coaxial grid and cathode lines are

replaced by parallel-wire lines and if the shorting plungers of the coaxial lines are replaced by the adjustable shorting bars on the parallel-wire lines. The plate supply voltage is applied between the E_{bb} terminal and center tap of the heater transformer. Grid-bias voltage results from grid current in R_g and is stored on the stray capacitance between the grid line and the shielded enclosure.

Two modes of oscillation of nearly the same frequency are possible in this circuit. In the desired mode, the two tubes operate in push-pull—the voltage of each transmission line is balanced with respect to the shield, and the r-f voltage at one cathode (or one grid) is positive when that of the other electrode is negative. In the undesired mode, the two tubes operate in parallel—the two cathodes and the two conductors of the cathode line form the center conductor of a transmission line for which the shielding box is the outer conductor. The grid line functions similarly, and the oscillator becomes equivalent to the single-tube circuit of Fig. 7.

The undesired mode can be prevented in several ways. The transmission lines used by the parallel mode extend to the ends of the shielding enclosure, whereas the lines used by the push-pull mode terminate at the adjustable shorting bars. Furthermore, the grid line terminates in an open circuit (except for R_g) for the parallel mode and in a short circuit for the push-pull circuit. Thus it is possible to adjust line lengths so that the relation $f_k < f_g$ (see Art. 4) required for oscillation is satisfied for the push-pull mode but not for the parallel mode. Damping introduced by resistors R_g and R_{bb} may also be used to suppress the parallel mode, because these resistors carry r-f currents if oscillations in the parallel mode occur but do not appear in the r-f circuit for the balanced push-pull oscillations. If R_{bb} is of appreciable magnitude, it is shunted by the r-f choke L_{bb} to permit application of the full supply voltage E_{bb} to the tubes.

Because of the balanced nature of the push-pull circuit, the r-f potentials of the plates and the center points of the shorting bars for the grid and cathode lines are equal. Thus these three points are connected in the r-f equivalent circuit, and the circuits of Fig. 8 apply exactly to each tube in Fig. 9. The effects of tuning adjustments are therefore the same as described in Art. 4.

A ring oscillator may be formed by combining any number of push-pull oscillators into a single circuit—hence the even number of tubes in the ring circuit. The tubes are arranged symmetrically in a circle, and interconnections of the push-pull components are made in such a way that all tubes operate at the same frequency and in a definite phase relation. In the desired mode of a ring oscillator, a phase reversal occurs between each tube and the next along the circle; that is, each tube may be considered to operate in push-pull with either tube adjacent to it.

As an example of a ring oscillator, consider the circuit of Fig. 10, which employs triodes of conventional design but otherwise is an extension of the push-pull plate-return circuit of Fig. 9 to four-tube ring form. The cathode resonator in Fig. 9 is the four-tube version of the parallel-wire cathode line in Fig. 8, but instead of grid lines, the four small inductances L_1, L_2, L_3, and L_4 are employed. Each inductance is made adjustable by means of a tuning disk (rotating a disk toward the plane perpendicular

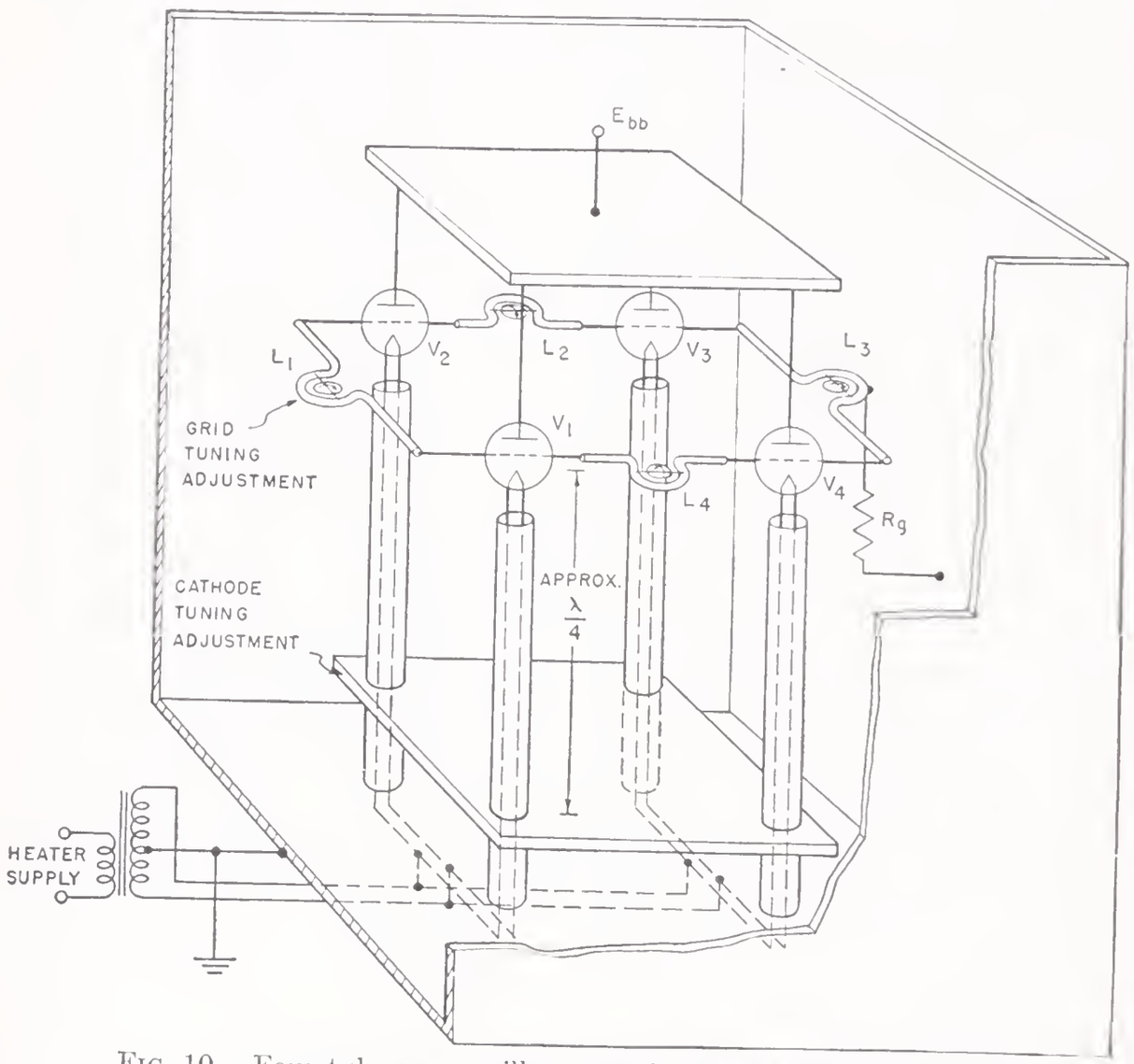


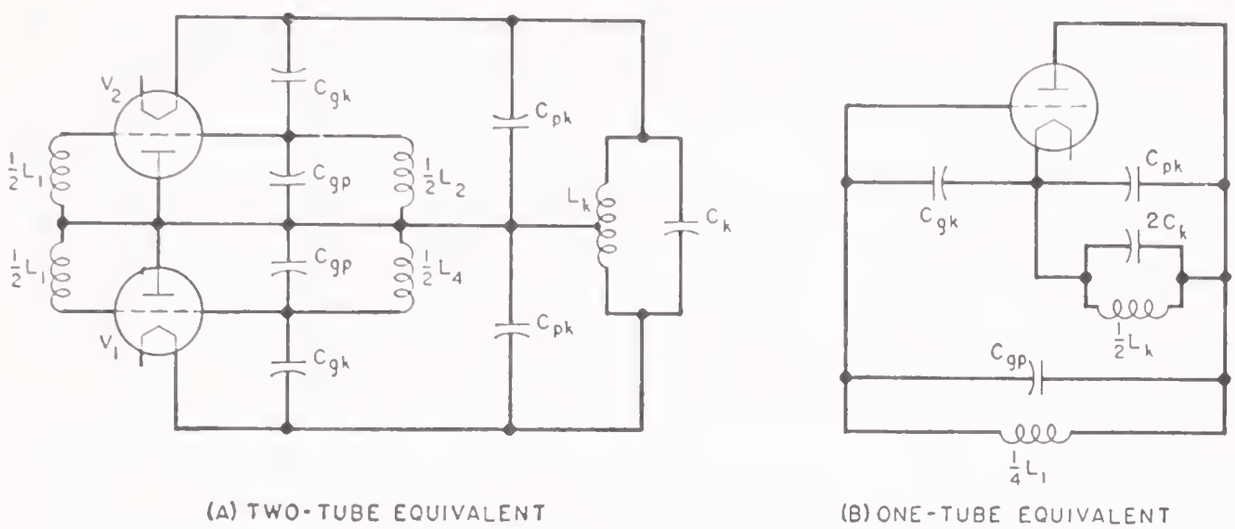
FIG. 10. Four-tube ring oscillator employing plate-return circuit.

to the magnetic flux decreases the inductance), and each tube is provided with two grid leads for convenient connection to the inductances.

For operation in the desired mode, the cathode circuit of any pair of adjacent tubes may be considered to be a parallel-wire line formed by the hollow cathode conductors for these tubes and to have a characteristic resistance determined in part by the cathode line of the other pair of tubes. The cathode-line parameters for tubes V_1 and V_2, for example, are those which would apply if a conducting plane extended vertically upward from the cathode shorting plate midway between the circuit of V_1 and V_2 and that of V_3 and V_4. The reason these parameters apply

is that a positive voltage at any point in the circuit of V_1 is matched by a negative voltage symmetrically located in the circuit of V_4 and similarly voltages in the V_2 circuit are matched by voltages of opposite polarity in the V_3 circuit. As a result of this symmetry, the voltage at every point of the plane midway between the two pairs of tubes is zero, and the imaginary conductor could be inserted without altering the field.

An r-f equivalent circuit for V_1 and V_2, when operation is in the desired mode, is drawn in Fig. 11*A*. The two grids are joined on one side by the inductance L_1; and because the center point of L_1 is maintained at the potential of the anode shorting plane by the balanced nature of the oscillations, a connection is made between the midtap of L_1 and the anodes of



(A) TWO-TUBE EQUIVALENT (B) ONE-TUBE EQUIVALENT

FIG. 11. R-f equivalent circuits for oscillator of Fig. 10.

V_1 and V_2. The balanced oscillations also maintain the center point of L_2 and L_4 at plate potential, and thus $\frac{1}{2}L_2$ and $\frac{1}{2}L_4$ appear in Fig. 11*A*. The resonant circuit L_k, C_k represents the cathode line of V_1 and V_2 (with the imaginary conducting plane taken into account) and is tuned to the frequency for quarter-wave resonance of that line.

The circuit of Fig. 11*A* may be simplified to the one-tube circuit of Fig. 11*B*, in which the inductance value $\frac{1}{4}L_1$ applies if the four grid inductances are equal. This final equivalent circuit is the same as that of the single-tube and push-pull plate-return oscillators and indicates that frequency is controlled by the grid-circuit disks and that the magnitude of the feedback voltage depends upon the position of the cathode-line shorting plate.

One of the undesired modes of oscillation may be called push-pull parallel operation. Tubes V_1 and V_4, for example, may operate in phase as one unit of a push-pull oscillator with V_2 and V_3 operating in phase as the other unit. In such a mode, inductances L_2 and L_4 would carry no current, and the major change of the equivalent circuit of Fig. 11*B* would be the increase of $\frac{1}{4}L_1$ to $\frac{1}{2}L_1$. Adjustments can be made so that this change lowers the resonant frequency sufficiently to make the cathode

circuit inductive and thus prevent the tubes from sustaining this mode. Another undesired mode is parallel operation of all four tubes. In this mode no currents flow in any of the grid inductances, and the grid-plate impedance of Fig. 11*B* becomes essentially capacitance C_{gp}. Resonance then requires an inductive plate-cathode impedance and oscillations cannot be sustained.

Of particular interest is the similarity of the multiconductor resonators employed in ring oscillators to the cavity resonators used in magnetrons. The natural modes of magnetron resonators are considered in detail in Art. 6, Chap. XI, and the principles developed there apply also to the ring-type resonators of uhf oscillators.

6. Analysis of Grid-return Circuits. The designation *plate return* applied to the oscillators of Arts. 4 and 5 is part of a general classification scheme for triode circuits illustrated by Fig. 12. The cathode-return

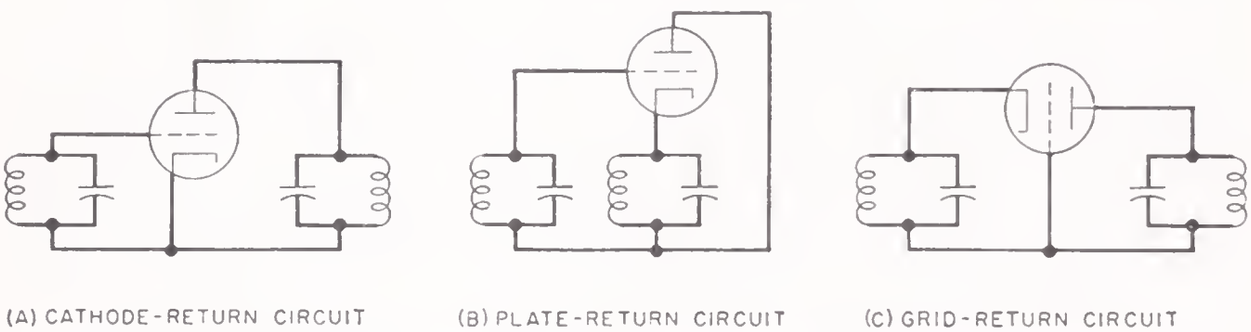


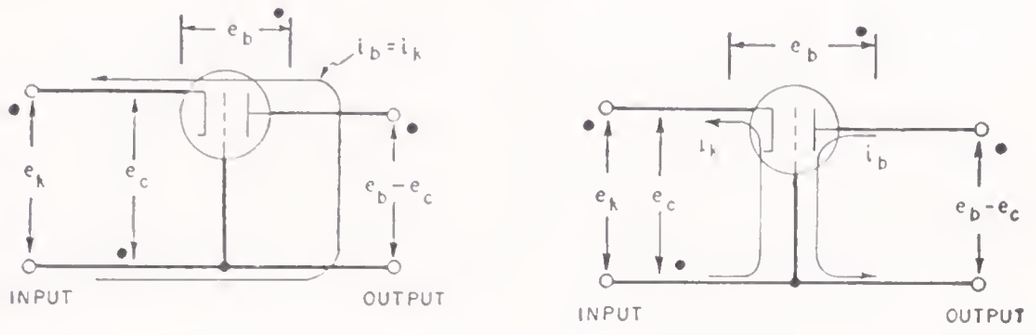
Fig. 12. Circuits illustrating three classifications of triode connections.

connection is employed in most low-frequency circuits, and the plate-return connection has a low-frequency application in the cathode follower. The grid-return connection is used occasionally at low frequencies (for example, the normally off tube of the cathode-coupled multivibrator of Fig. 32*A*, p. 100 in Chap. II, is a grid-return or grounded-grid amplifier). At the highest useful frequencies of triodes, the grid-return circuit is of great importance both for amplifiers (see Art. 25, Chap. VI, and Art. 15, Chap. XII) and for oscillators. The basic reason for the importance of the grid-return circuit in uhf applications is the physical position of the grid between cathode and plate in a triode. Grid-return circuits employing disk-seal tubes can be made in which the only coupling between input and output cavities, other than that of the electron beam, is the very small coupling produced by the penetration of electric fields through the grid mesh.

The entire structure of vacuum-tube analysis is ordinarily based upon the cathode-return circuit. For example, voltages measured with respect to the cathode are used in plotting tube characteristics, and linear equivalent circuits represent the effect of a grid-cathode input signal upon a plate-cathode output circuit. This procedure is entirely logical in the study of low-frequency circuits because of the prime importance

of cathode-return circuits. At high frequencies, however, grid-return circuits are of major interest, and a method of thinking based upon the grid as a dividing element between the input and output circuits is preferable. One advantage of the grid-return approach is that the high-frequency behavior of triodes, which appears very complicated on the cathode-return basis, can be expressed much more simply in terms of an input grid-cathode region and an output grid-plate region. Of interest also is the very considerable similarity between the operation of a triode considered on the grid-return basis and the operation of a klystron (see Art. 1, Chap. XI).

Voltages and currents in a grid-return circuit are indicated in Fig. 13. If the operating frequency is low, and if a sufficiently negative grid-bias voltage is employed, grid current is negligible, and the circuit of Fig. 13*A*



(A) LOW-FREQUENCY, NEGATIVE-GRID CIRCUIT

(B) GENERAL CIRCUIT

FIG. 13. Voltages and currents in grid-return circuit.

applies. The input voltage is the cathode voltage e_k, which is the negative of the grid-cathode voltage e_c. The output voltage is the plate-to-grid voltage $e_b - e_c$. The plate current i_b is equal to the cathode current i_k and to the negative of the current delivered by the input source to the cathode, or by the plate to the load.

The paths of currents i_b and i_k can be separated as in Fig. 13*B*. For low-frequency, negative-grid operation, the current $i_b - i_k$ in the grid lead is zero, so that Figs. 13*A* and *B* are equivalent. If, however, the grid voltage may be positive, or if capacitance currents or transit-time grid loading are important, grid current cannot be ignored, i_b and i_k are not equal, and the circuit of Fig. 13*B* must be employed. If the tube in Fig. 13*B* has a high amplification factor, the fields in the grid-plate region have little influence upon the electrons between cathode and grid. Thus the cathode-grid section of the tube is very nearly equivalent to a diode, and the source of voltage e_k may be considered to supply the current $-i_k$, determined by this diode, with little influence from the plate voltage. Most of the electrons reaching the grid plane pass through the grid and travel to the plate. In the grid-plate region, the electrons are accelerated

by the field of the direct component of plate-to-grid voltage and decelerated by the field of the alternating component of this voltage (see Art. 2). Thus the electrons serve to change d-c energy of the plate supply into a-c energy, which may be stored in a resonator connected between plate and grid. An equivalent statement of output-circuit relationships is that the current i_b produces the output voltage by flowing through the impedance of a resonator in the output circuit.

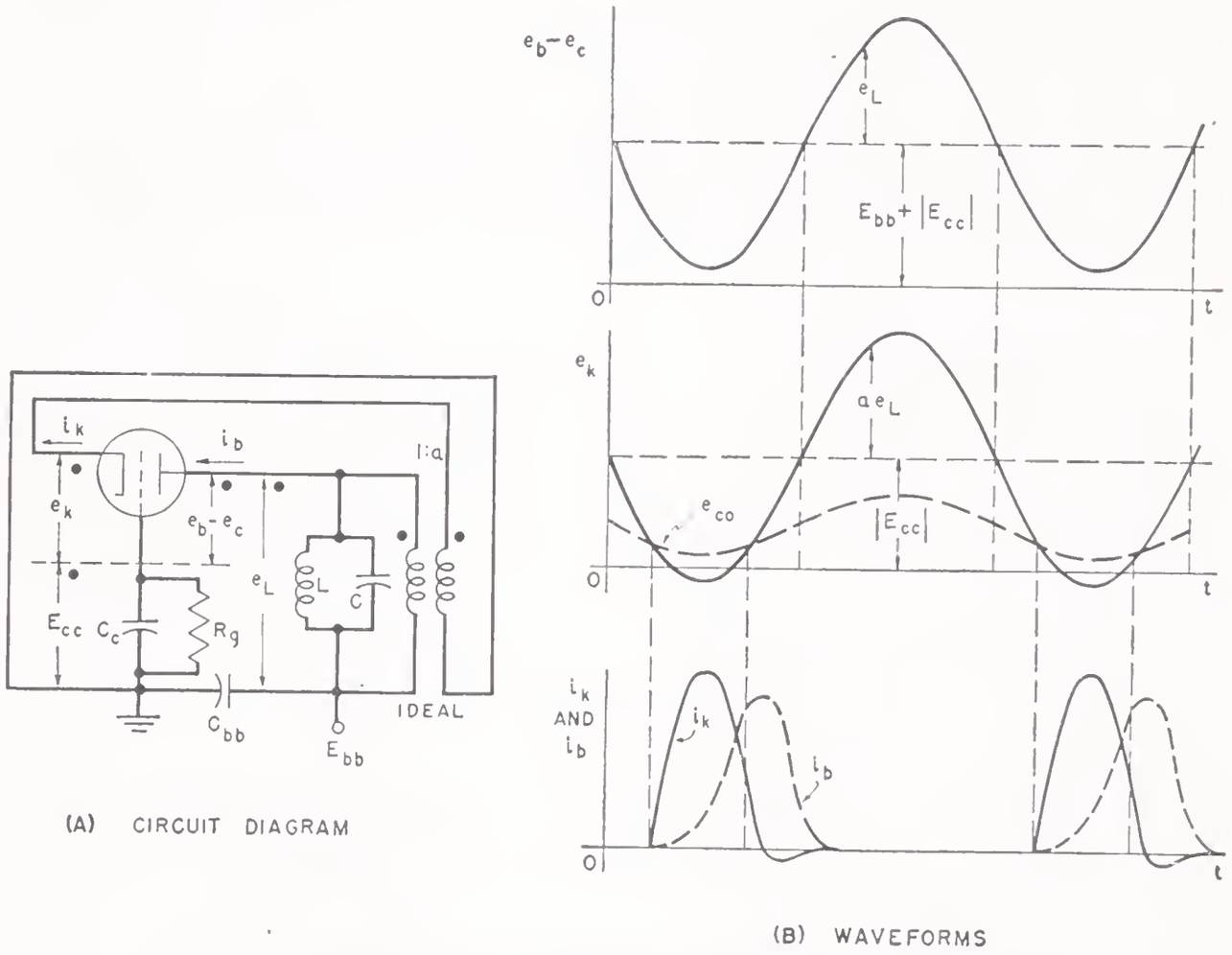


FIG. 14. Idealized circuit illustrating principles of grid-return circuit.

In Fig. 14*A* is drawn an idealized circuit that illustrates the grid-return approach to oscillators in the same way that the tuned-plate oscillator of Fig. 3 illustrates the cathode-return approach. Observe that the direct plate voltage is applied between plate and cathode in the usual way and that a conventional grid-leak-bias circuit is employed. The r-f equivalent circuit, however, takes the grid-return form of Fig. 13 because of by-pass capacitors C_c and C_{bb}. The waveform of plate-grid voltage $e_b - e_c$ in Fig. 14*B* is the sum of the plate supply voltage, the bias voltage across R_g, and the voltage e_L produced by oscillating energy in the *L-C* resonator. A portion ae_L of the resonator voltage is applied by the ideal transformer to the cathode circuit *in phase* with the plate-grid voltage. This feedback voltage added to the bias voltage results in the e_k waveform shown.

Electrons leave the cathode during the time that the voltage e_k is less than the cutoff voltage (or the grid voltage is above cutoff). These electrons, in traversing the cathode-grid space, give rise to the cathode-current pulse i_k in Fig. 14*B*, and in traversing the grid-plate region at a slightly later time, the electrons produce the plate-current pulse i_b. The reversed current at the end of the i_k pulse is the result of electrons in the grid-cathode space being returned to the cathode when the grid-cathode field reverses. (See the discussion of back bombardment in Art. 1.) The plate-current pulse occurs when the r-f component of $e_b - e_c$ is negative and therefore contributes energy to the resonator oscillations. The efficiency of the circuit could be improved if the phase of the a-c component of e_k could be shifted to provide an i_b pulse at the time of minimum $e_b - e_c$. The waveforms of Fig. 14*B* apply only for an oscillation frequency well below the ultimate limit of the tube. When a tube is operated at very nearly its maximum frequency, the oscillations are so weak and the spreading of current pulses because of transit time is so great that Class C conditions are not obtained.

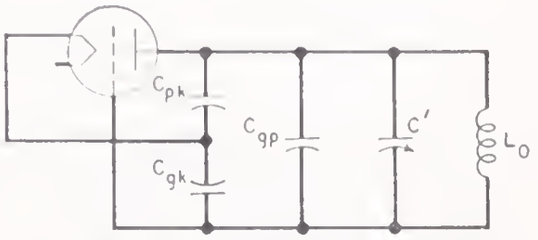


FIG. 15. R-f equivalent circuit of ultra-audion oscillator drawn in grid-return form.

The possibility of oscillation and the efficiency of operation of this circuit depend upon the transformer ratio a. In Fig. 14*A*, a must exceed a certain value in order for the plate-current pulses to be of sufficient magnitude to sustain oscillations. Moderate increase of a above this critical value increases the output of the oscillator, but excessive values of a increase the relative magnitude of the power consumed by the input circuit and decrease the r-f output voltage. Oscillation is impossible if a is increased to or beyond unity—a step-down voltage ratio is necessary in order that the cathode-current pulses in the transformer secondary winding may be equivalent to smaller pulses in the primary winding and thus may consume less power from the resonator than the plate-current pulses add to it.

The cathode-return method of analysis is not restricted to cathode-return circuit arrangements; neither is the grid-return approach restricted to circuits in grid-return form. Either method of analysis can be applied to any oscillator circuit, but one or the other may be easier to use, depending upon the circuit form. The grid-return analysis is well-adapted to most uhf circuits, particularly those employed with disk-seal tubes.

To illustrate the equivalence of the grid-return and cathode-return analyses, the ultra-audion equivalent circuit of Fig. 5*B* is redrawn in Fig. 15 in an arrangement appropriate to the grid-return analysis. The

output-circuit resonator in this diagram is the same as the frequency-determining resonator of Fig. 5*B*, and the resonant-transformer effect of the capacitance voltage divider that supplies feedback voltages to the cathode replaces the ideal transformer of Fig. 14*A*. Observe that the polarity of the returned voltage is correct and that the required step-down voltage ratio is provided.

7. Grid-return Oscillators.[1] One form of tuner used with disk-seal tubes is illustrated by the lighthouse-tube oscillator of Fig. 16. The tuner employs coaxial-line cathode-grid and anode-grid cavities, the cathode line surrounding the anode line so that the entire assembly comprises three concentric cylinders fitted to the tube at one end. The end of the grid cylinder at the tube is partly cut away (see figure) to make the

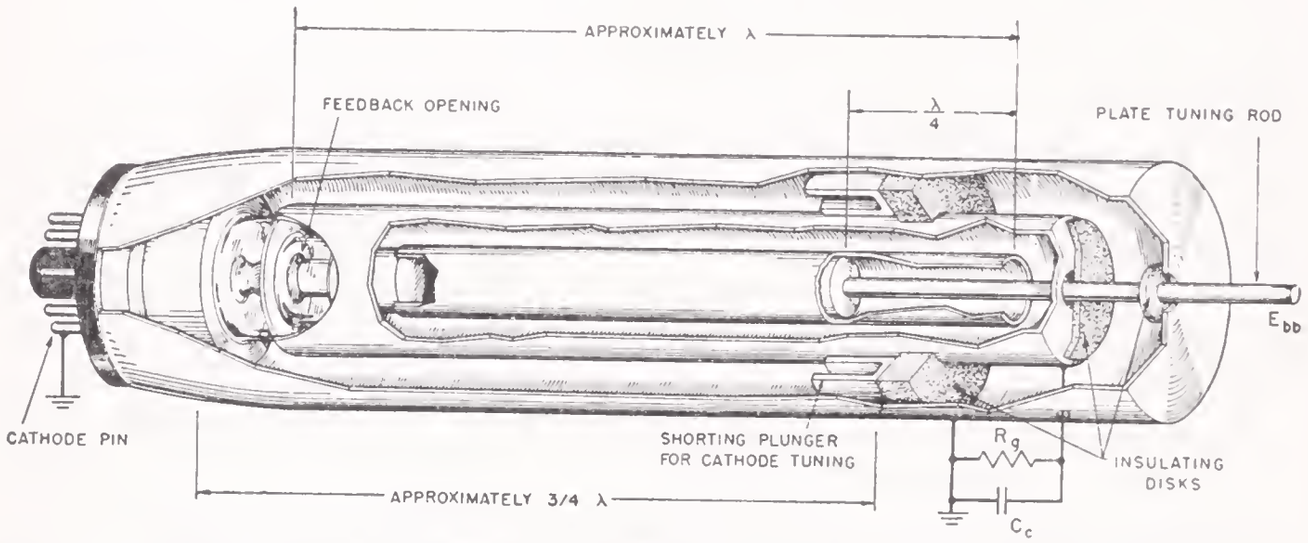


FIG. 16. Lighthouse tube in coaxial tuner employing cathode line folded over plate line

cathode-plate capacitance greater than the very small value produced by the fringing of fields through the grid mesh. A very similar tuner, but one with no openings in the grid cylinder, is employed in r-f amplifiers (Art. 15, Chap. XII).

Because of the high frequency at which this oscillator operates—about 3,000 Mcps—mechanical considerations require use of tuning lines greater than a quarter wave in length. The cathode line is approximately three-quarters of a wavelength, measured from the tube to the adjustable short-circuiting plunger, and the plate line is nearly a full wavelength from tube to the end of the plate cylinder. A noncontacting form of short-circuiting plunger (see Art. 5, Chap. IX) is used in the cathode line to permit grid-bias voltage to develop across R_g. The re-entrant quarter-wave line inside the plate cylinder serves as an r-f choke in the lead to

[1] A. M. GUREWITSCH and J. R. WHINNERY, "Microwave Oscillators Using Disk-seal Tubes," *Proc. I.R.E.*, **35** (May, 1947), 462–473; G. LEHMAN, "Very-high-frequency Triode Oscillator and Amplifier Circuits," *Elec. Comm.*, **25** (March, 1948), 50–61; D. R. HAMILTON, J. K. KNIPP, and J. B. H. KUPER, *Klystrons and Microwave Triodes*, 170–198 (McGraw-Hill Book Company, Inc., New York, 1948).

the plate voltage supply, and the open-end construction of the plate line assists in insulating the high plate voltage from the grid circuit. Tuning of the plate line is accomplished by sliding the plate cylinder over the plate terminal of the tube.

Both the short-circuited three-quarter-wave cathode line and the open-circuited full-wave plate line have input impedances representable by parallel L-C circuits (Art. 8, Chap. IX). Therefore, the r-f equivalent circuit may be drawn as in Fig. 17. In this figure C_{pk} represents the sum of the tube capacitance and the capacitance effect of the openings in the grid cylinder.

Because the voltage divider that supplies the alternating cathode voltage must not cause a polarity reversal, the net reactance of L_k, C_k, and C_{gk} in parallel must be capacitive. That is, the cathode line loaded by C_{gk} must resonate at a frequency f_k lower than the frequency of oscillation f_0. The frequency of oscillation is the resonant frequency of L_p in parallel with C_p, C_{gp} and the input capacitance of the voltage divider. Therefore f_0 is less than f_p, the resonant frequency of the plate line loaded by capacitance C_{gp}. Oscillations can occur only if line lengths are adjusted so that f_k is less than f_p.

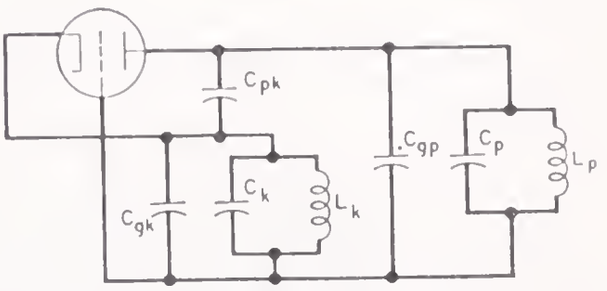


FIG. 17. R-f equivalent circuit for oscillator of Fig. 16.

The effects of tuning adjustments are very similar to those obtained in the plate-return circuit. The plate-line length is the basic frequency control, because it determines the resonant frequency of L_p and C_p in Fig. 17. The cathode-line length determines the cathode-grid reactance in the equivalent circuit and therefore controls the amplitude of the feedback voltage. Oscillations cease if the cathode line is made so short that the cathode-grid reactance becomes inductive or if the line is made so long that the reactance, though capacitive, becomes too small to provide a feedback voltage of adequate amplitude.

An adjustable element for the control of phase of the feedback voltage is not provided in the oscillator of Fig. 16. Nevertheless, some control of phase may be exercised in the design of the oscillator through adjustment of the size of the openings in the tube end of the grid cylinder. Because the coupling reactance (see Art. 7, Chap. IX) through which the resonator of Fig. 17 supplies current to the cathode is capacitive, the phase of the cathode voltage is shifted somewhat ahead of the phase of the plate-grid voltage. As may be seen from Fig. 14B, this shift is in the proper direction for obtaining maximum efficiency when transit time is appreciable. The amount of the phase shift is increased if the size of

the grid-cylinder openings, and therefore the magnitude of capacitance C_{pk}, is decreased. To obtain the maximum possible efficiency, the proper combination of opening size and cathode-line length must be utilized.

Because the oscillator of Fig. 16 employs multiple quarter-wave resonances in the tuning lines, the possibility of parasitic oscillation in lower frequency modes must be considered. Such oscillations may be expected to occur unless specific provision for suppressing them is made, because the tube is a more effective amplifier at the lower frequencies. Lower frequency ranges in which the impedance of the grid-plate resonator in Fig. 16 is appreciable are those for which the plate line—measured from the tube around the end of the plate cylinder to the short circuit inside the re-entrant line—is in approximate one-quarter-wave or three-quarter-wave resonance. These resonances occur at frequencies of about

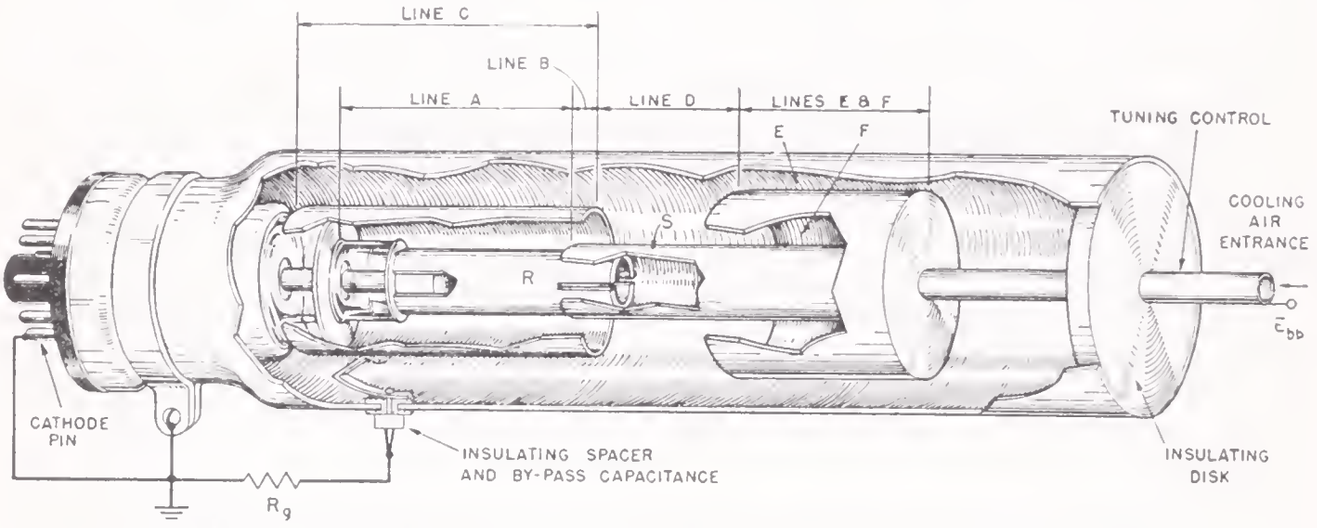


FIG. 18. Lighthouse-tube oscillator employing transmission-line feedback.

one-fifth and three-fifths of the desired frequency and correspond to cathode-line lengths of roughly $^3/_{20}$ and $^9/_{20}$ wave. For the first of these lengths, the cathode-resonator impedance is inductive, and for the second it is nearly a short circuit. Thus the feedback required to sustain the unwanted modes is not provided.

The circuit of Fig. 17 could equally well be drawn in grounded-cathode form and thought of as a Colpitts circuit. There is no great advantage in using the grid-return rather than the cathode-return approach in the study of this circuit. For another lighthouse-tube oscillator having the form sketched in Fig. 18, however, the grid-return approach offers a considerable advantage in simplicity.

The d-c circuits for the oscillators of Figs. 16 and 18 are the same, but the r-f circuits differ, principally in the paths provided for feedback from the plate-grid resonator to the cathode. Only one tuning adjustment is provided in the oscillator of Fig. 18, and by means of it the frequency can be adjusted over a range of about 10 per cent. Moving the tuning control causes the plate sleeve S to slide with respect to the plate rod R.

The r-f equivalent circuit for Fig. 18 is somewhat complicated by the several junctions and changes of impedance in the coaxial-line-tuner structure. A circuit that does not represent all the discontinuity effects in detail, yet is satisfactory for a qualitative explanation of the mechanism of oscillation, is drawn in Fig. 19. In this, the grid-plate capacitance appears as C_{gp} but the plate-cathode capacitance, which is too small to have appreciable effect, is omitted. The grid-cathode capacitance, augmented by the capacitance of the gap between the grid cylinder and the shell of the tube, is referred through the 1:1 ideal transformer and designated C_{gk} in the equivalent circuit.

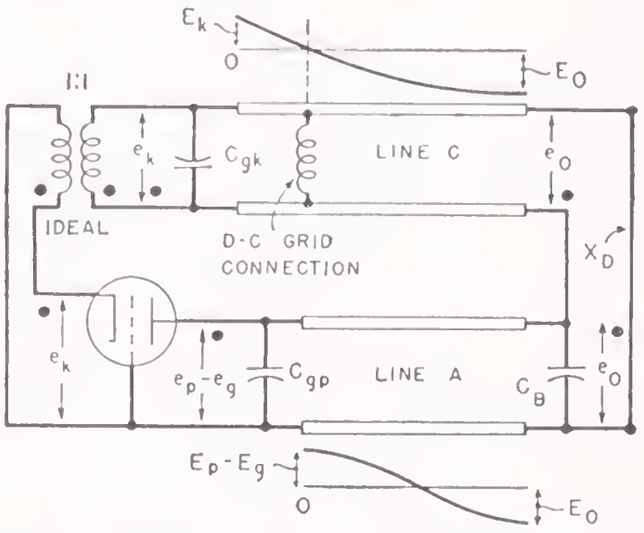


FIG. 19. R-f equivalent circuit for oscillator of Fig. 18.

Line A in Fig. 19 represents the similarly marked coaxial line inside the grid cylinder in Fig. 18. Line B in Fig. 18 is short and has a low characteristic impedance; thus it appears as a capacitance C_B in Fig. 19. The conductors for line C in Fig. 18 are the grid and cathode cylinders, whereas for line D the conductors are the cathode cylinder and the plate sleeve. The series junction of lines B, C, and D is represented at the right-hand side of the equivalent circuit.

The input reactance X_D for line D is shown in Fig. 19 as a short circuit because the line is a quarter wave in length and is terminated by an open circuit. The terminating open circuit results from the series combination of a short circuit (at the input to line E) and an open circuit (at the input to line F). Lines E and F are each a quarter-wave long and make up a choke that is effective in confining the r-f fields to the oscillator cavity without short-circuiting the high direct voltage between plate and cathode conductors.

Line C connects the junction of lines B, C, and D to the cathode-grid terminals of the tube. In the equivalent circuit, this connection is made through an ideal transformer in order to avoid showing the upper conductor of line C in parallel with the lower conductor of line A. In the actual oscillator, the two lines are entirely independent; the inner surface of the grid cylinder may be considered one conductor of line A and the outer surface one conductor of line C. Because the transformer is ideal and has a 1:1 ratio, it is exactly equivalent to a direct connection, except for its effect in emphasizing the independence of the two lines.

Line C is somewhat less than a half wave in length and resonates with capacitor C_{gk}. As a result of this resonance, the standing-wave pattern

indicated in Fig. 19 is obtained. The voltage e_0 has an rms value E_0 greater than E_k, the rms value of the cathode voltage, and the resistance at the cathode end of the line that represents power consumed by the tube is equivalent to a larger resistance at the input terminals of the line. This larger resistance loads the resonant system comprising line A and capacitors C_{gp} and C_B. Because of these capacitors, the resonant length of line A is much less than a half wave, and the standing-wave pattern is as shown—E_0 and $E_p - E_g$ may be assumed approximately equal. Observe that the d-c grid connection is made to the point of minimum voltage on line C; thus the inductance of the connecting finger has a negligible effect upon the r-f circuit, and little r-f energy can escape to R_g.

The two transmission lines and three capacitors of Fig. 19 are the frequency-determining resonator of the oscillator. Adjusting the tuning rod controls the resonator frequency by means of two effects. The first effect is a change in the length of line D, which change causes the short circuit of Fig. 19 to become an inductive or capacitive reactance. The second effect is a change in the length of line A and an opposite change in the length of line B (corresponding to a change in the capacitance C_B). The first effect is much less important than the second, because the reactance X_D is located at a voltage maximum in the standing-wave pattern of lines A and C; thus the current in X_D is small.

The frequency-determining resonator also functions as a resonant transformer to supply the voltage e_k to the cathode in proper relation to the plate-grid voltage $e_b - e_c$. The standing-wave diagrams of Fig. 19 indicate an in-phase relation of e_k to $e_b - e_c$ and a step-down voltage ratio, as required for oscillation. The standing-wave patterns are drawn for total reflection at the cathode and thus ignore the power input to the cathode-grid section of the tube. Actually, the standing-wave ratio for the cathode line is only moderately high not only because of the triode input power but also because a load, not shown in Figs. 18 and 19, is coupled to line C. Thus a leading phase for e_k can be obtained by appropriate adjustment of line lengths. In order to provide optimum phase and amplitude of the feedback voltage, two parameters of the tuner (for example, the length of the grid cylinder and the spacing of the grid-cylinder-to-cathode gap) must be adjusted. Through correct adjustment of the various tuner dimensions, nearly optimum feedback can be had throughout the 10 per cent tuning range of the oscillator. Thus single-control adjustment of frequency is accomplished.

8. Loading of Oscillators. An oscillator has been presented in the previous articles as a resonator in which oscillations are sustained by means of a vacuum tube. In order for the oscillator to be useful, a load (the antenna of a radar set, for example) must be connected to the resonator and supplied with high-frequency power. In steady-state opera-

tion, the power output of the tube must equal the sum of the load power and the loss in the resonator. If the power supplied to the resonator exceeds the power consumed, energy is added to the oscillating fields during each cycle of operation, and the amplitude of oscillation grows until an increase of power consumed or decrease of power supplied causes equilibrium to be established. Similarly, if the power supplied is less than the power consumed, the stored energy of the resonator decreases. Either equilibrium is reached at a low amplitude of oscillation, or the oscillations cease altogether.

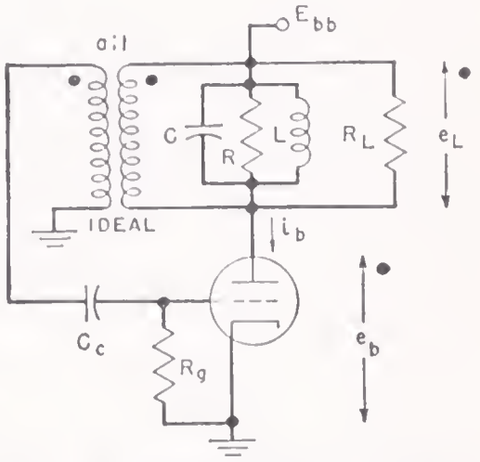


FIG. 20. Tuned-plate oscillator with load.

In order that an oscillator may provide a maximum of useful load power, not only must the phase and amplitude of the feedback voltage be properly adjusted and the grid-bias voltage correctly chosen, but also the correct load resistance must be presented to the tube by the resonator. The principles of oscillator loading may be explained on the cathode-return basis by reference to the tuned-plate oscillator of Fig. 3, p. 704. The circuit of Fig. 3*A* is redrawn in Fig. 20, with the resistors R and R_L

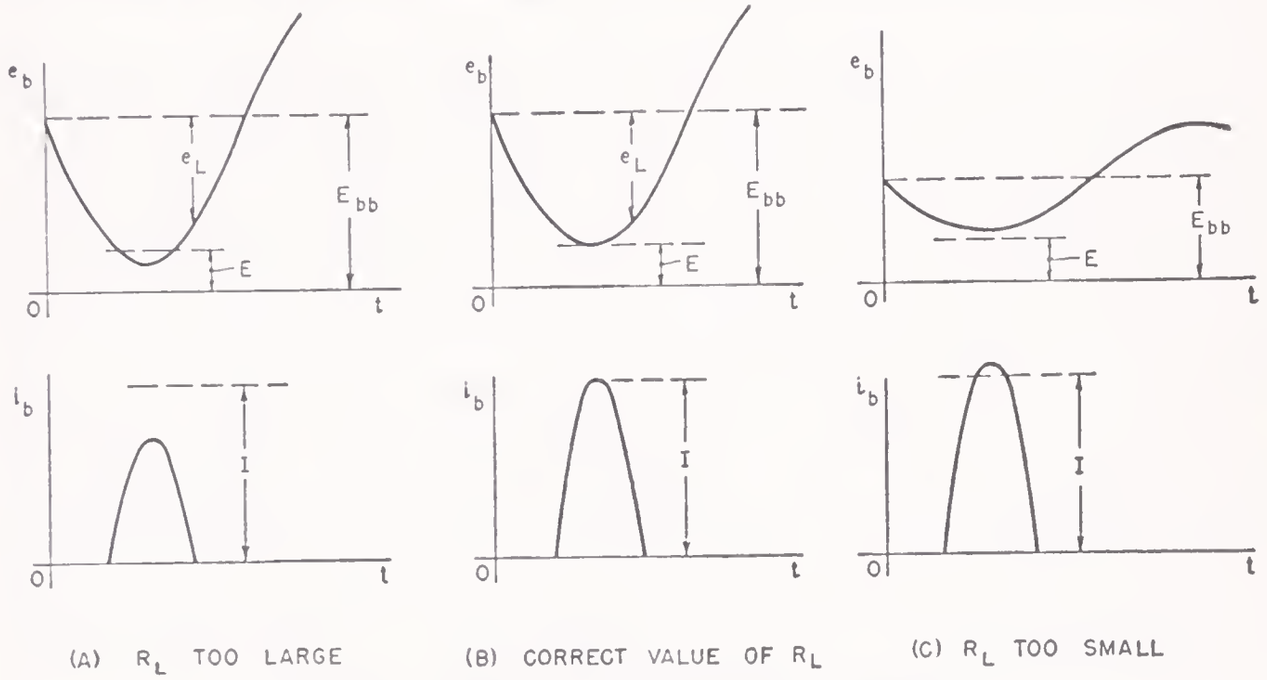


FIG. 21. Waveforms for oscillator of Fig. 20.

added to represent, respectively, the loss resistance of the resonator and a load resistor connected to the resonator terminals.

Waveforms of plate voltage and plate current, corresponding to the waveforms of Fig. 3*B*, are drawn in Fig. 21 for three different values of the load resistance R_L. The waveforms of Fig. 21*B* are obtained if R_L is adjusted so that the combination of R and R_L is the plate-load resistance

into which the tube can deliver the greatest amount of power. In this condition of operation, the plate-voltage wave reaches a minimum value E of the order of one to two times the peak positive value of the grid-voltage wave (see the waveform of e_c in Fig. 3*B*), and the plate-current wave reaches the maximum value I. The current limit may be determined by the emission capabilities of the cathode or, for a given value of E, by the allowable power dissipation at the anode.

If R_L is made greater than the value of Fig. 21*B*, the first effect is an increase in the amplitude of the oscillating resonator voltage. The increase results from the momentary excess of the power supplied by the tube over the power consumed by the high-resistance load. However, the increase of resonator voltage is small, as indicated in Fig. 21*A*, because the decrease of the minimum plate voltage below E decreases the amplitude of the plate-current pulse and thus decreases the power supplied by the tube. The feedback voltage or the grid-bias voltage may be readjusted in an attempt to increase the peak grid voltage and thus obtain large plate current with small plate voltage. Such a procedure, however, greatly increases the power consumed by the grid circuit and may decrease the plate current because of the interception of electrons by the positive grid.

If R_L is made less than the value of Fig. 21*B*, the immediate effect is an increase in power delivered by the resonator to R_L and a decrease of the resonator voltage, as indicated in Fig. 21*C*. The consequent increase of the minimum plate voltage above E increases the peak value of the plate-current pulse, provided the current is not emission limited, and the changes of both current and voltage contribute to an increased plate dissipation. In order that the tube may be operated within its power rating, the plate supply voltage must be decreased, as in Fig. 21*C*, and the net effect is a loss of power output caused by operation with nearly the same current wave as in Fig. 21*B* but a smaller resonator voltage.

The over-all conclusion to be derived from Fig. 21 is that more or less definite limits E and I exist upon the minimum value of plate voltage and the maximum value of plate current. Maximum power with respect to adjustment of R_L is delivered by the tube for the particular choice of R_L that permits operation with the plate voltage and plate current each reaching its limiting value. Observe that this value of R_L is determined by the tube ratings, *not* by a consideration of matching R_L to an internal impedance of the oscillator. (Contrast the statements made for linear generators in Art. 14 of Chap. VI.)

The effects of load-resistance adjustment are essentially the same if a grid-return circuit rather than a cathode-return circuit is considered. In the oscillator of Fig. 22 (which is the circuit of Fig. 14*A*, p. 714, with loss resistance R and load resistor R_L added), for example, too large a

value for R_l makes the output power small because the voltage limit requires operating with small plate-current pulses, and too small a value for R_l limits the power because the resonator voltage produced by the maximum plate current is small. The correct value of R_L is somewhat greater for this oscillator than for the tuned-plate circuit because the sum of the grid and plate voltages, instead of the plate voltage alone as in the tuned-plate oscillator, appears across R_L in Fig. 22.

The discussion of oscillator loading has thus far been concerned with the power delivered by the tube to the resonator. The division of this power between the loss resistor R and the load resistor R_L in Figs. 20 and 22 is also important. Values of R very much greater than R_L are desired in order that nearly all the tube power may be received by the load. An equivalent statement is that the unloaded Q of the resonator, $Q = R/\sqrt{L/C}$, should be very high relative to the loaded Q,

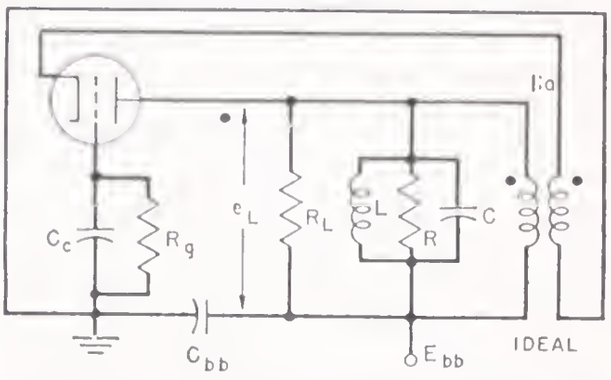


FIG. 22. Grid-return oscillator with load.

$$Q_l = \frac{RR_L/(R + R_L)}{\sqrt{L/C}}$$

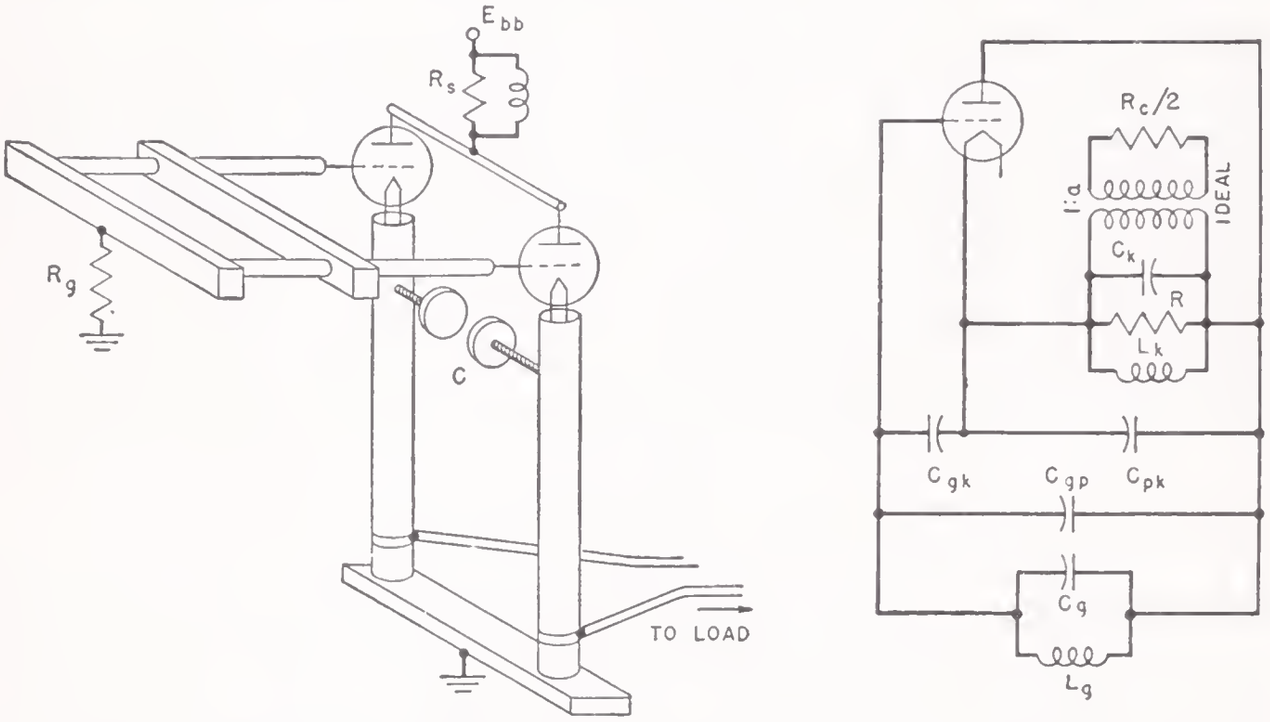
To provide a high value of R or a high ratio of Q to Q_l, both Q and the impedance-level parameter $\sqrt{L/C}$ should be a maximum. Large resonator-conductor surfaces made of low-resistivity material contribute to a high value of Q, and the minimizing of miscellaneous effects such as power loss along the power-supply lead and in high-resistance contacts between resonator parts is especially important. The impedance-level parameter tends to decrease with increase of the operating frequency. Minimum values for interelectrode capacitances in uhf triodes (see Art. 1) are important if $\sqrt{L/C}$ is to be made as high as possible.

9. Output Circuits for Uhf Oscillators. An oscillator output circuit may utilize any one of the resonator coupling devices discussed in Art. 9, 13, 14, and 16 of Chap. IX; that is, a transmission-line junction, a coupling loop, or a coupling probe may be used to connect the resonator of the oscillator to the transmission line that delivers power to the load. An adjustable element in the coupling device is usually provided in order that the load resistance for the tube may be set to the optimum value. A second adjustable element sometimes provided permits the reactance as well as the resistance contributed to the resonator by the output circuit to be controlled.

The transmission-line junction is a particularly simple coupling device but can be used only if no direct voltage exists between the load and the

oscillator line to which the output connection is made. In the plate-return oscillator, the cathode line is often at ground potential, as in the push-pull circuit of Fig. 23*A*, and the output line can therefore be connected as shown. The output taps on the cathode line may be moved to adjust the loading, and capacitor C provides a feedback adjustment equivalent to the adjustment of cathode-line length in the plate-return circuit of Fig. 9, p. 712.

The equivalent circuit of Fig. 23*B* is obtained by adding the loss resistance R and the output-circuit elements to the circuit of Fig. 8*A*,



(A) SEMIPICTORIAL REPRESENTATION (B) R-f EQUIVALENT CIRCUIT FOR ONE TUBE

FIG. 23. Plate-return oscillator with output line connected to cathode line.

p. 711. The output line is assumed to be terminated in its characteristic resistance R_c, so that its input impedance is also R_c, and the resistance $R_c/2$ effective in the circuit of each tube appears in Fig. 23*B*. The ideal transformer represents the resonant-transformer effect of the cathode line and has the ratio a equal to the ratio of the voltage between output taps to the voltage between cathodes. If the taps are moved nearly to the shorting bar, a is small, the resistance $R_c/2a^2$ effective in the plate-cathode circuit is large, and the lightly loaded oscillator delivers little power. By moving the taps toward the tubes, $R_c/2a^2$ is decreased and the output power increased. Too high a position for the taps makes the load resistance so small that the power obtainable without overloading the tubes is reduced.

The characteristic impedance of output lines is of the order of 100 ohms, and the optimum plate load resistance per tube is usually about 1,000 to 10,000 ohms. The transformation ratio a must therefore lie in the

neighborhood of L_{10}, and an output-tap position rather near the shorting bar is required. An inductive coupling reactance of appreciable magnitude (see Art. 9, Chap. IX) may therefore be present in the connection between the cathode line and the output line. This coupling reactance is ignored in Fig. 23*B*; its effect is considered in Art. 10.

If the oscillator lines from which output power is obtained are at a high direct voltage, as are the plate lines in Fig. 24, inductive coupling may be employed to deliver r-f power to the output line. The oscillator of Fig. 24 is a four-tube ring oscillator employing a grid-return connection. In the desired mode of oscillation (Art. 5), the grids may all be considered to be at the same r-f potential as the center points of the plate and cathode lines because the grids of tubes V_1 and V_4 (which tubes operate

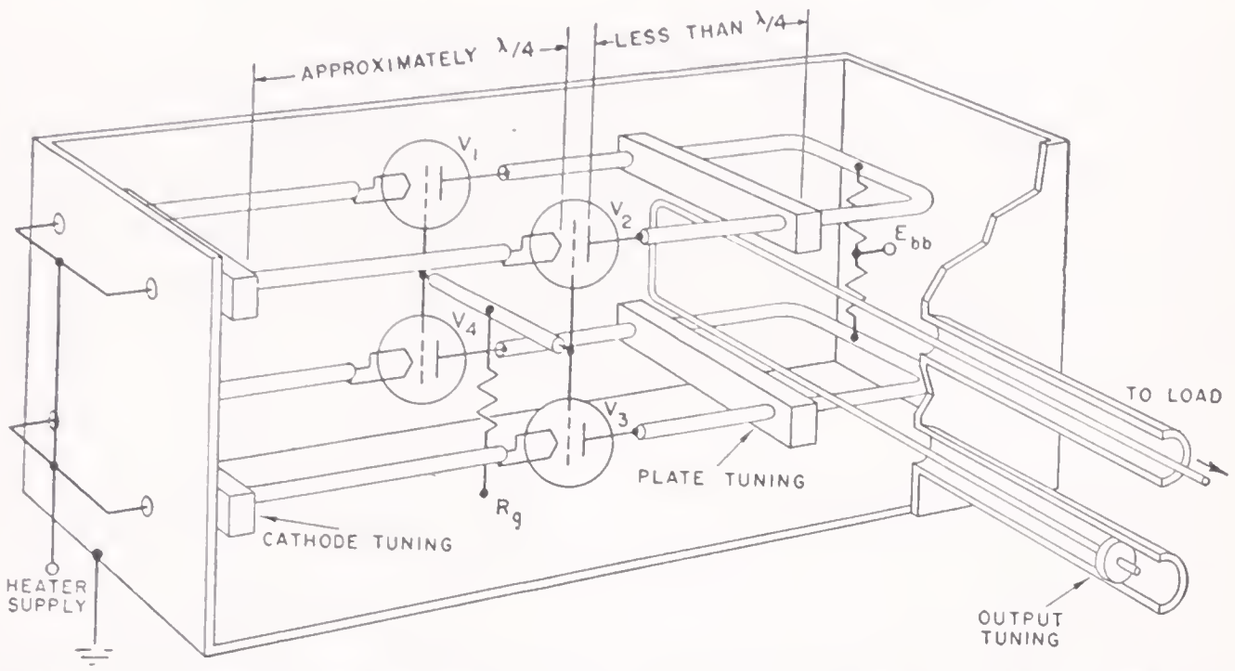


FIG. 24. Grid-return oscillator with output circuit inductively coupled to plate lines.

out of phase) are joined by a short circuit and the grids of tubes V_2 and V_3 are connected similarly. The longer grid-circuit connection joining tubes V_1 and V_4 to tubes V_2 and V_3 enters the r-f circuit only for a wrong mode of oscillation in which V_1 and V_4 in parallel form one unit of a push-pull oscillator, with V_2 and V_3 as the other unit. This mode is prevented by the reduction of feedback voltage occasioned by the additional grid-circuit inductance. In the desired mode, r-f currents flow in opposite directions through the two plate-line shorting bars, and thus the two currents send magnetic flux through the coupling loop in the same direction.

A one-tube r-f equivalent circuit for Fig. 24 is drawn in Fig. 25. This circuit differs from that of Fig. 17 only in the addition of the loss resistance R and the output-circuit elements. As explained in Art. 9, Chap. IX, the coupling loop can be represented by the connection to the plate-line resonator through the ideal transformer and the coupling inductance

L_c. The ratio a of the transformer is the ratio of the voltage induced in the output loop by plate-line currents to one-half the plate-to-plate voltage of the oscillator; hence a is increased if the coupling loop is altered to link more magnetic flux. Resistor R_c represents the impedance of the output line, and X_1 is the reactance of the coaxial-line stub containing the output-tuning adjustment.

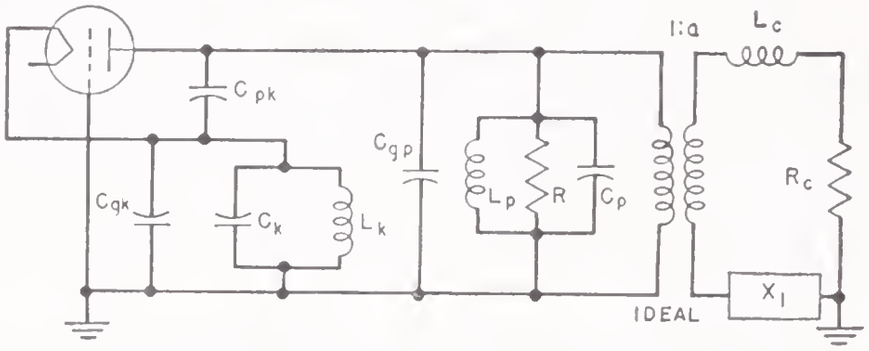


FIG. 25. R-f equivalent circuit for one tube of oscillator of Fig. 24.

The output-tuning control is normally adjusted so that X_1 is a capacitive reactance in series resonance with the inductive coupling reactance. Thus the total impedance in the secondary circuit of the ideal transformer is R_c, and R_c/a^2 is the resistance presented by the transformer to the plate-grid circuit of an oscillator tube. The oscillator loading can therefore be controlled through adjustment of the loop; increasing the mag-

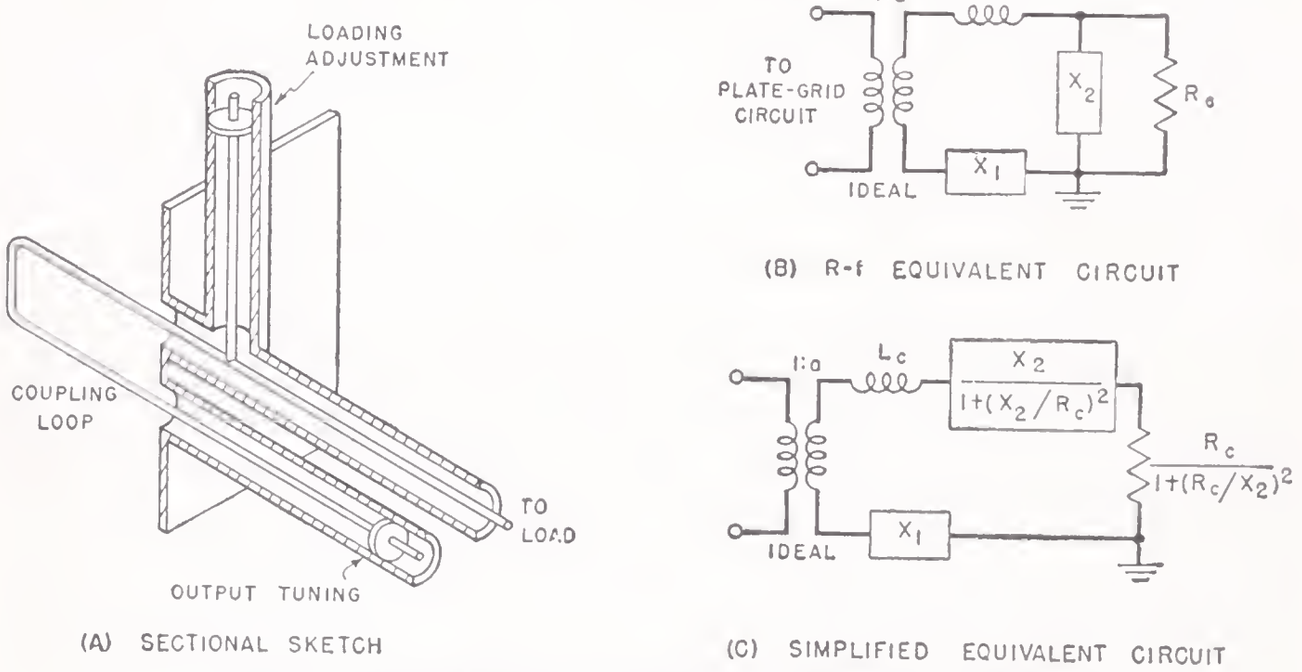


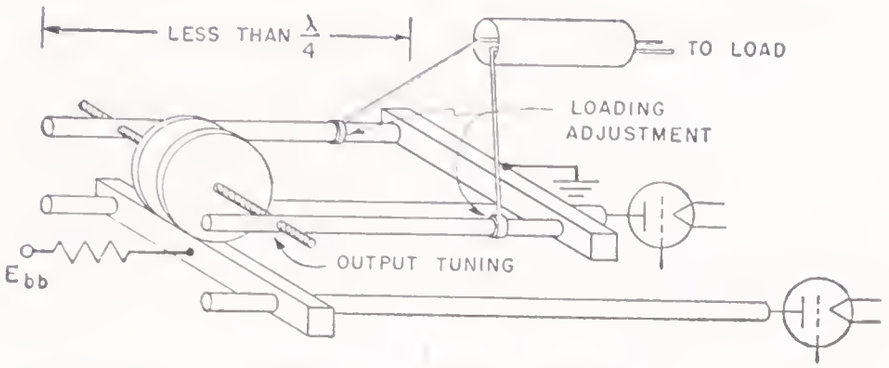
FIG. 26. Alternate output circuit for oscillator of Fig. 24.

netic coupling increases the ratio a and reduces the resistance R_c/a^2 effective in the plate circuit.

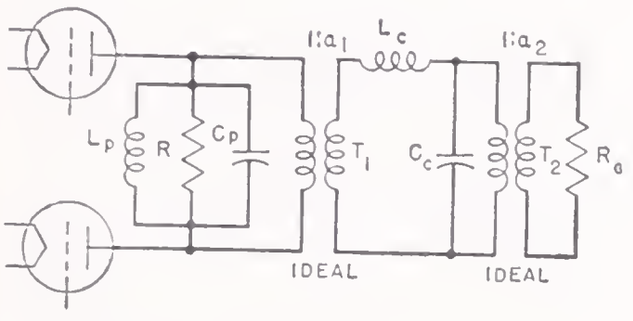
Sometimes the position of the coupling loop is fixed, and an adjustable shunt impedance at the junction of the coupling loop and output line is used to control the loading of the oscillator. A modification of this kind for the output circuit of Fig. 24 is indicated in Fig. 26A, and an equiva-

lent-circuit representation of the modification is indicated in Fig. 26*B*. At the oscillator frequency, the parallel combination of output-line resistance R_c and loading-stub reactance X_2 can be changed to the series resistance-reactance combination of Fig. 26*C*. By adjustment of X_2, the series resistance can be made to equal the value required by the oscillator, and by adjustment of X_1, series resonance of the output circuit may be obtained.

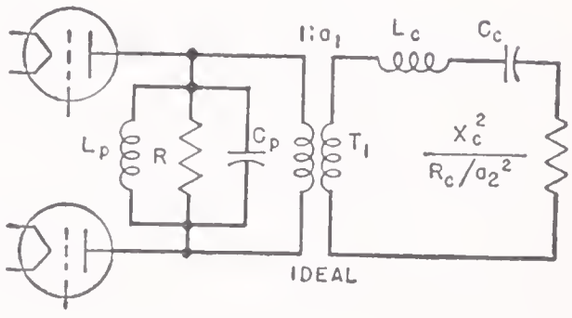
A further modification of this output circuit consists in omitting the output tuning stub from Fig. 26 and adding a second shunt stub to the output line. The double-stub tuner (see Art. 15, Chap. VII) thus formed can be adjusted so that the impedance faced by the coupling loop



(A) SEMIPICTORIAL REPRESENTATION



(B) R-f EQUIVALENT CIRCUIT



(C) SIMPLIFIED EQUIVALENT CIRCUIT

FIG. 27. Coupling loop and plate line of push-pull oscillator.

contains the capacitive reactance required for output-circuit resonance and the resistance required for proper loading of the oscillator. Note that the required load impedance for the coupling loop can be provided even though the output line is not terminated in its characteristic resistance. In order to minimize line losses and to avoid rapid change of input impedance with frequency, however, the final load should always be adjusted for minimum standing-wave ratio on the output line, and proper oscillator operation should be secured by adjustments at the oscillator end of the line.

A different form of coupling loop suitable for use with a push-pull oscillator is illustrated in Fig. 27*A*. A transmission-line section coupled inductively to the oscillator plate line is tuned to resonance by the

capacitance disks, and the output line is connected to adjustable taps on the coupling line. In the equivalent circuit of Fig. 27*B*, the plate line is represented by L_p, C_p, and R, and loop coupling to this line is represented by the ideal transformer T_1 and the coupling inductance L_c. The transformer ratio a_1 is determined by the ratio of voltage induced in the coupling line to the plate-to-plate voltage of the oscillator. The tuning capacitance for the coupling line appears as C_c, and the transformation ratio between the voltage across C_c and the voltage between the output taps is represented by the ideal transformer T_2 of ratio a_2.

The simplified equivalent circuit in Fig. 27*C* is obtained from Fig. 27*B* as follows: Transformer T_2 and resistor R_c are replaced by a resistance R_c/a^2, and the parallel combination of R_c/a^2 and C_c is changed to a series *R-C* circuit equivalent to the parallel circuit at the oscillator frequency. For normal positions of the output taps, R_c/a^2 is much greater than X_c, the reactance of C_c, and thus the approximate series-circuit parameters indicated in Fig. 27*C* may be employed. Figure 27*C* indicates that the coupling line may be tuned for series resonance of L_c and C_c, and the plate-to-plate load impedance for the oscillator then becomes $(a_2/a_1)^2X_c^2/R_c$.

The presence of both ratios a_1 and a_2 in the expression for the load resistance corresponds to the fact that the oscillator loading may be changed by moving the coupling line toward or away from the plate load or by adjusting the taps on the coupling line. As might be expected, moving the coupling line toward the plate line loads the oscillator more heavily—the increased voltage induced in the coupling line increases the value of a_1 and thus reduces the load resistance. The effect of the tap adjustment, however, is opposite to that obtained in the direct-coupled system of Fig. 23*A*. *If the output taps are moved closer to the shorting bar, the oscillator is loaded more heavily*—the ratio a_2 is decreased, and the load resistance $(a_2/a_1)^2X_c^2/R_c$ is therefore made smaller. If the output line is disconnected from the coupling line, the secondary of transformer T_1 in Figs. 27*B* and *C* is connected to the high-Q series-resonant circuit comprising L_c and C_c, and the plate-to-plate load resistance is nearly a short circuit.

The correct value of load resistance can be obtained for many different combinations of coupling-line position and output-tap position. The optimum adjustment is obtained by locating the coupling line as near as possible to the plate line and placing the taps relatively far from the shorting bar at the position that yields the correct load resistance. One reason for using this adjustment is that the loaded Q of the coupling line is made small, thus reducing the circulating currents and loss in this line. A second reason concerns the effect of load on frequency and is discussed in Art. 10. Observe that the coupling loop and two coaxial-line stubs of

Fig. 26*A* may be thought of as a distorted half-wave line with tap connections for the load. Thus the output systems of Figs. 26*A* and 27*A* are very similar, and a high-Q output circuit is to be avoided in Fig. 26 for the same reasons as in Fig. 27.

In single-tube oscillators employing coaxial-line tuners, coupling probes are frequently employed to connect the output line to the resonator of the oscillator. The probe does not short-circuit the direct voltage that may exist between cylinders of a coaxial-line tuner, and the penetration of the probe into the resonator may be adjusted to control the oscillator loading.

As an example of the use of a coupling probe, a portion of the tuner for the plate-return oscillator of Fig. 7, p. 710, is redrawn in Fig. 28 with a coupling probe added to the cathode line. The probe penetration is

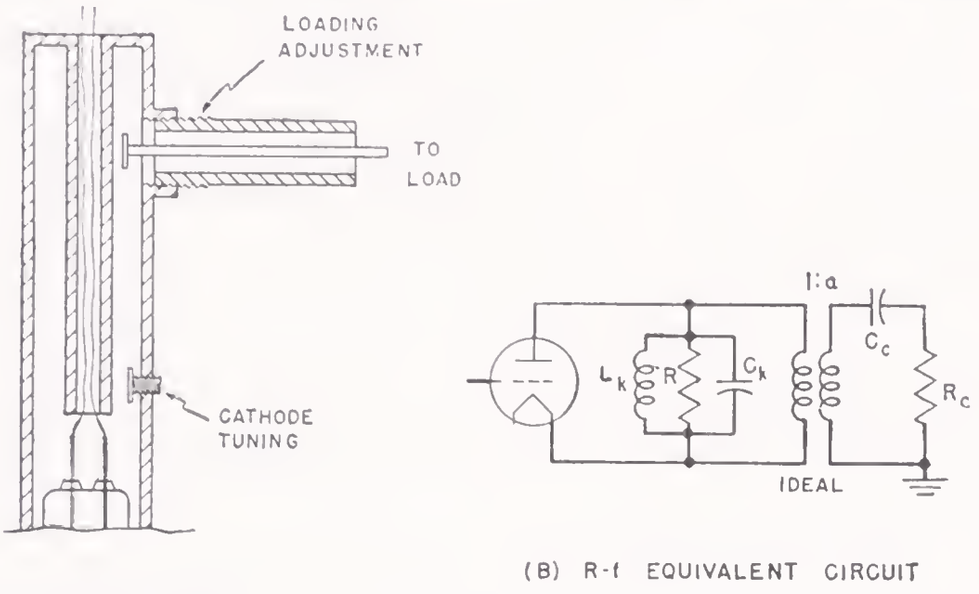


FIG. 28. Coupling probe for single-tube plate-return oscillator.

adjusted by means of the screw fitting that connects the output line to the cathode line. The equivalent circuit in Fig. 28*B* for the cathode line and output circuit is based upon the probe equivalent developed in Art. 16 of Chap. IX. The ideal-transformer ratio a is the ratio of the probe voltage to the plate-to-cathode voltage of the resonator, and C_c is the coupling capacitance of the probe. The ratio a may be determined by calculating the probe voltage from the effective length of the probe and the electric-field intensity. Alternatively, a may be thought of as the product of two factors: (1) the ratio of the voltage between resonator conductors at the probe position to the plate-to-cathode voltage and (2) the fraction of the voltage between conductors applied to the probe by the capacitance-voltage-divider effect of the probe.

If the probe capacitance is large, the reactance of C_c may be negligible, so that (see Fig. 28*B*) the plate-to-cathode load resistance is R_c/a^2 in parallel with the loss resistance R. Increasing the penetration of the probe increases a, decreases R_c/a^2, and loads the oscillator more heavily.

Frequently, the reactance of C_c is not negligible, and a shunt stub is added to the output line. By adjustment of this stub, resistor R_c in Fig. 28*B* is changed to an impedance having an inductive component that resonates with capacitor C_c. If the probe is long or the terminating disk large, series resonance between the probe capacitance and the inductance of the probe stem may reduce the net coupling reactance.

Oscillator output coupling frequently requires that balanced parallel-wire oscillator lines be joined to unbalanced coaxial output lines, or vice versa. A method of employing a coupling loop in such a connection is illustrated by Fig. 24. Other arrangements for balanced-to-unbalanced couplings are discussed in Art. 9, Chap. XII.

10. Effects of Load on Frequency of Oscillation. When adjustments are made in output-coupling circuits of an oscillator, it is usually found

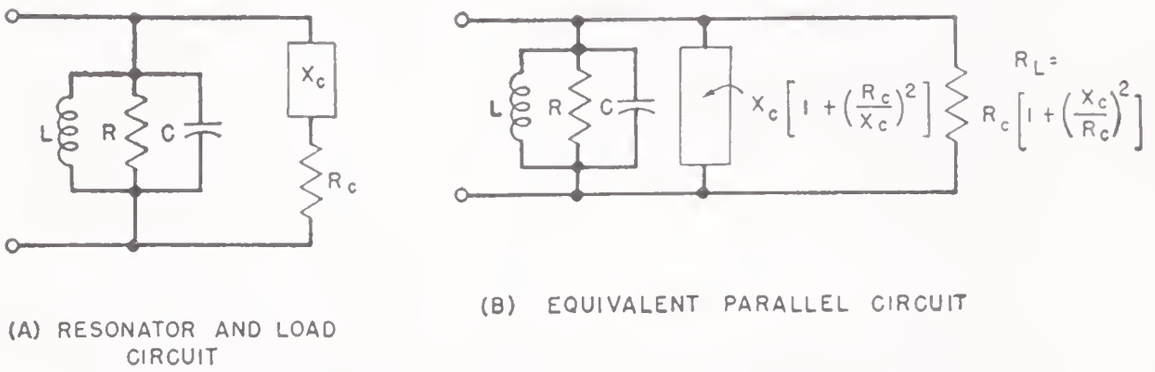


FIG. 29. Oscillator output circuit with untuned coupling reactance.

that not only the loading but also the frequency of the oscillator is changed. Sometimes the frequency varies smoothly as the adjustment is made; at other times a sudden jump in the frequency occurs. Similar effects are obtained if the load impedance changes (as may occur if the oscillator is a radar transmitter because of antenna rotation). Smooth variation of frequency with change of load impedance is sometimes called *frequency pulling*, and sudden jumps of frequency are known as *frequency splitting*. The detailed effects of load on frequency depend upon the particular output circuit used, but the principles involved may be explained by reference to two simple circuits.

The first of these circuits is drawn in Fig. 29*A*. The parallel combination of L, C, and R represents the resonator of the oscillator, and R_c represents a *resistance* load connected to the oscillator through the coupling reactance X_c. Such a circuit may arise, as explained in Art. 9, if the load is connected to an untuned coupling loop or probe or to taps near the shorting bar of an oscillator line. The effect of the load is most easily seen if the series branch of Fig. 29*A* is replaced by an equivalent parallel circuit (at the oscillator frequency), as in Fig. 29*B*. The influence of load upon frequency is evident from Fig. 29*B*, because the load-circuit reactance $X_c[1 + (R_c/X_c)^2]$ in parallel with L and C affects the resonant fre-

quency of the over-all circuit, and the magnitude of this reactance is altered when R_c is adjusted. Thus it is seen that, whereas frequency shifts may be expected to result from changes of load reactance, they may also be caused by load resistance changes if coupling reactances are appreciable.

Another important conclusion derivable from Fig. 29*B* is that the load resistor

$$R_L = R_c\left[1 + \left(\frac{X_c}{R_c}\right)^2\right]$$

effective across the resonator can never, by adjustment of R_c, be made less than $2X_c$. If R_c is very large, R_L is nearly equal to R_c and is also large. Reducing R_c reduces R_L to a minimum value $2X_c$ for $R_c = X_c$, and further reduction of R_c causes R_L to increase; R_L becomes an open circuit for $R_c = 0$. Because of this effect, a large coupling reactance may prevent loading an oscillator sufficiently to permit the tubes to develop the full power output. To avoid this limitation, most of the output circuits of Art. 9 include tuning devices that may be adjusted for zero net reactance in series with the load.

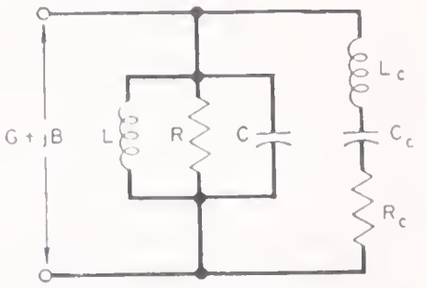


FIG. 30. Oscillator output circuit with tuned coupling reactance.

Use of series output tuning results in the circuit of Fig. 30, which is the second of the two circuits to be discussed. Either L_c or C_c may represent unavoidable coupling reactance, the other element being introduced deliberately to permit series tuning. If L_c and C_c resonate at f_0, the resonant frequency of L and C, it might appear that oscillation would occur at this frequency and that loading in any desired degree could be accomplished by adjustment of R_c. Unfortunately, an attempt to load the oscillator heavily may result in an oscillation frequency different from f_0 and in reduced loading at the new frequency.

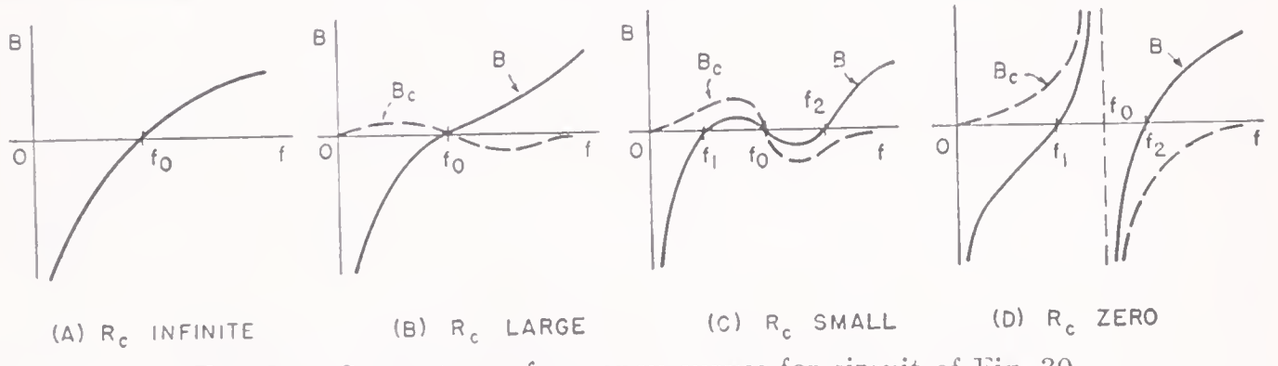


FIG. 31. Susceptance-frequency curves for circuit of Fig. 30.

Figure 31 is drawn to explain this shift of frequency. The susceptance component B of the resonator-and-load-circuit admittance $G + jB$ (see Fig. 30) is plotted as a function of frequency in Fig. 31. Figure 31*A*, which applies if R_c is open-circuited, is the susceptance curve of the

parallel L-C resonator. If R_c is large, but not infinite, the series-resonant circuit comprising L_c, C_c, and R_c has a low Q and adds the small susceptance B_c indicated by the dashed curve in Fig. 31B to the parallel-circuit susceptance. Adding the B_c curve to the curve of Fig. 31A yields the solid curve of Fig. 31B. The curve of Fig. 31C applies for a smaller value of R_c. Susceptance B_c is greater than in Fig. 31B and has more effect upon the over-all curve. The solid curve of Fig. 31D, applicable for $R_c = 0$, shows that, when R_c is short-circuited, the resonator has two open-circuit modes of resonance (at f_1 and f_2) and one short-circuit mode (at f_0).

For the conditions of Figs. 31A and B, oscillation occurs at the frequency f_0, where $B = 0$, corresponding to the parallel resonance of L and C. Similarly, for the conditions of Figs. 31C and D, oscillations can occur at either of the frequencies f_1 and f_2, where the B, f curve indicates conditions similar to parallel resonance. Frequency f_0 corresponds to resonance of a series L-C circuit and is *not* a possible oscillation frequency.

As R_c is decreased from larger values in an attempt to increase the load applied to the oscillator, the frequency of oscillation changes from f_0 to either f_1 or f_2 when the susceptance curve changes from the form of Fig. 31B to that of Fig. 31C. The frequency shift prevents the desired increase of oscillator load because L_c and C_c are nonresonant at the new frequency and contribute a reactance in series with R_c. Operation in a condition corresponding to Fig. 31C is undesirable because small changes in the tuning of the resonator or output circuits can cause sudden jumps between frequencies f_1 and f_2.

In the interest of frequency stability, not only should the conditions of Fig. 31C be avoided, but also the slope of the B, f curve at f_0 should be made as great as possible. In this way the frequency shift caused by small tube and circuit changes (which may be considered equivalent to susceptances added in parallel with the resonator) is minimized. A low impedance-level parameter $\sqrt{L/C}$ for the resonator is therefore desirable from this point of view, although high values of $\sqrt{L/C}$ are required to minimize resonator losses (see Art. 8). For a given resonator, the slope of the B, f curve near f_0 is greatest if the Q of the L_c-C_c-R_c circuit in Fig. 30 is made as low as possible. For these reasons, adjusting a tuning element for series resonance of the output circuit is not sufficient to ensure optimum oscillator operation; minimizing the coupling reactance is also necessary.

The detailed effects of load upon oscillator frequency differ greatly from one circuit to another. The B, f curves for the resonators may differ, as may the way in which reactances enter the output circuit. Furthermore, in many oscillators the load influences oscillator frequency through its effect upon the feedback voltage. Increasing the loading of

an oscillator, in general, reduces the magnitude of the feedback voltage and thus decreases the amplitude of the plate-current pulses. The load may in addition influence the phase of the feedback voltage. The plate-current pulses have been considered in previous articles to have only the effect of sustaining oscillations. However, if the pulses do not occur exactly at the time of minimum plate voltage, they have also an effect upon frequency, and thus the influence of load upon feedback has an indirect effect upon frequency. The effect of plate-current pulses on frequency is considered in detail in Art. 4, Chap. XI, in connection with electrical tuning of klystrons.

In order to specify the effects of load-impedance changes upon oscillator performance, it is necessary to select definite points in the oscillator circuit as the output terminals at which the load impedance is to be measured. For uhf oscillators, customary practice is to designate a particular cross section in the output transmission line as the oscillator output. The load impedance may then be specified as a per-unit impedance or as a reflection coefficient at the output cross section. For a given set of adjustments of the oscillator frequency, feedback, and loading controls, the frequency and power output may be measured for various standing-wave conditions in the output line. The results may be recorded as a *Rieke diagram* or plot on a Smith chart of frequency contours (lines of constant frequency) and power-output contours (lines of constant power output). For oscillators in which frequent adjustment of internal controls is impossible, the Rieke diagram forms a convenient method of determining the power supplied to various loads and of predicting frequency-pulling and frequency-splitting effects. The Rieke diagram is useful in the study of any high-frequency generator and, in fact, was first used with the magnetron. A magnetron Rieke diagram appears in Fig. 36, p. 802, and the diagram is discussed further in the magnetron chapter (see Art. 10, Chap. XI).

11. Grid-leak Bias in Oscillators. Thus far in this chapter only the steady-state condition in which r-f oscillations are sustained in the resonator of an oscillator have been considered. Also of interest are the processes by which oscillations start when a continuous-wave oscillator is turned on and by which oscillations repeatedly start and stop during operation of a pulsed oscillator. The bias voltage developed across the grid resistor R_g of any of the circuits in the preceding articles has an important effect upon the growth and decay of oscillation.

Consider, for example, the tuned-plate oscillator of Fig. 3*A*. Suppose that the plate supply voltage is applied while the grid is short-circuited to ground so that a steady plate current flows in the tube, as shown in Fig. 32. As soon as the grid short circuit is removed, oscillations begin, as indicated at time t_1 in the figure. Any transient disturbances which may

occur—noise currents, if nothing else—cause slight variations in the plate currents, which in turn produce a small oscillatory voltage across the resonant plate circuit. A portion of the plate-voltage oscillations is returned to the grid and causes additional variation of the plate current. If the feedback is sufficient, the oscillations grow in amplitude, as shown between times t_1 and t_2.

Since the grid voltage is positive during the peak of each cycle, grid current flows and produces a negative voltage across resistor R_g in Fig. 3*B*. Because of capacitor C_c, the bias voltage cannot build up immedi-

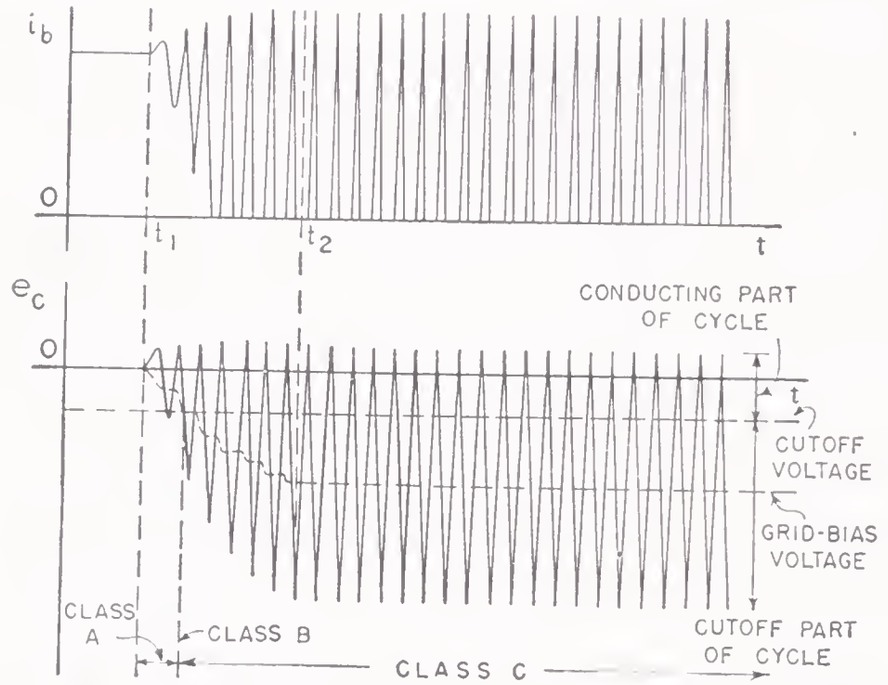


FIG. 32. Growth of oscillations when grid capacitor is small.

ately; instead, as the capacitor charges, the magnitude of the bias increases along the curve shown dotted in Fig. 32. If the capacitance of C_c is very small, as is assumed in Fig. 32, the bias voltage builds up nearly as fast as does the amplitude of oscillations.

Immediately after time t_1, the plate-current oscillations are sinusoidal. As the bias voltage grows, however, a point is reached at which the tube is cutoff for a portion of each cycle, and the plate current flows in short pulses. At the very beginning of the oscillation, the tube operation is Class A, but because of the bias voltage, the operation changes through Class B to Class C. Since the voltage increases until the charge placed on C_c by grid current at the peak of each cycle equals the charge lost through R_g during the cycle, the final bias voltage may be adjusted to the value for most efficient operation by proper choice of R_g.

The limiting values reached by the amplitude of oscillation and the bias voltage after the time t_2 are determined by the amount of power consumed by the load, the magnitude of the feedback voltage, and the value of the grid-leak resistance. At first the pulses of plate current are larger than are necessary to supply the output power, and the oscillations increase in

amplitude. The bias voltage also increases and causes the duration of the current pulses to become continually shorter. Therefore less energy is added to the oscillations, and at time t_2 the conducting periods are so reduced that the current pulses are just sufficient to maintain constant-amplitude oscillations.

The continuous steady oscillations that follow the initial growth period in Fig. 32 are in part the result of the assumption of a small size for capacitor C_c. If a large capacitor is employed, the oscillator may become *self-pulsed* or generate oscillations in short bursts, as indicated by the

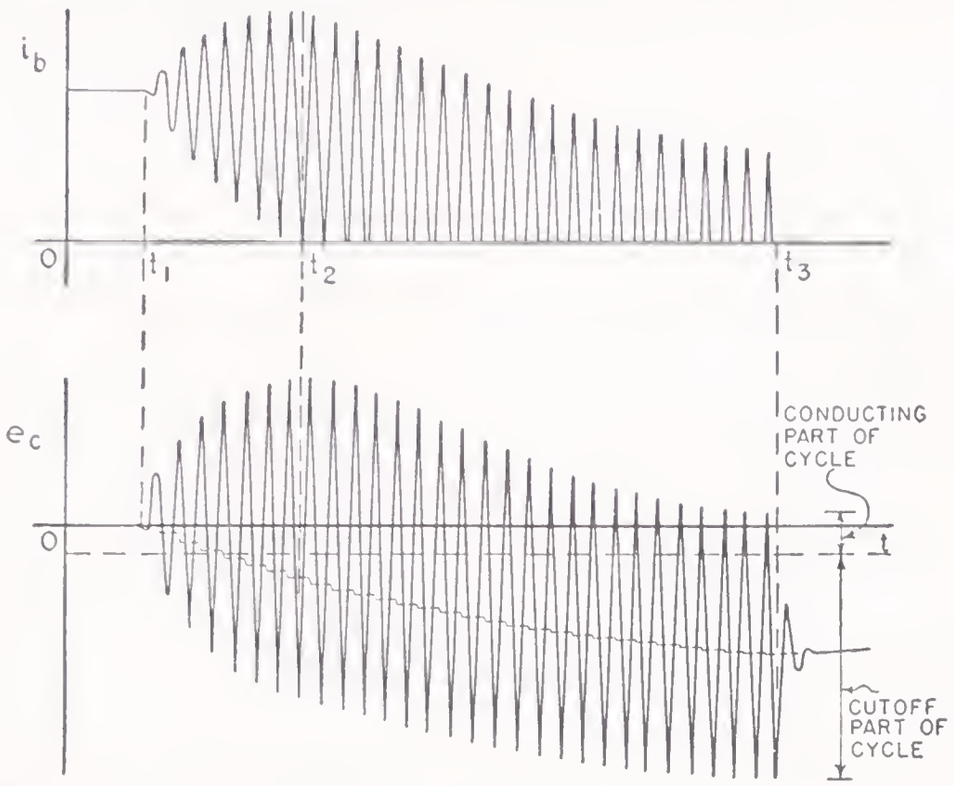


FIG. 33. Growth and decay of oscillations when grid capacitor is large.

waveforms of Fig. 33. Initial conditions of zero grid voltage and steady plate current are assumed exactly as before, and oscillations begin at time t_1 and grow in the same way. The amplitude of oscillations increases much faster than the bias voltage, however, because the large capacitor C_c cannot be charged quickly. Therefore, in Fig. 33, the oscillations reach final amplitude (at time t_2) before the grid-bias voltage has built up appreciably.

After time t_2, grid current continues to flow and to charge C_c. Therefore, the grid wave is a series of oscillations about an axis that moves slowly downward. The plate-current wave becomes a series of pulses, and the duration and amplitude of these pulses decrease as the bias voltage goes farther below cutoff.

As the plate-current pulses decrease, a time (indicated as t_3 in Fig. 33) is reached at which the pulses are not sufficient to supply the power taken from the resonator by the load. Then the amplitude of oscillation decreases. Since the grid-bias voltage can change only slowly, the

decrease of oscillation amplitude causes the conducting period to become an even smaller fraction of the cutoff period in each r-f cycle. Therefore, the tubes supply still less power, and the oscillations quickly die away altogether. The grid voltage then consists only of the voltage stored on the grid capacitor, and the oscillator tubes remain cutoff until the capacitor discharges through resistor R_g.

Large values of C_c can cause self-pulsing only if the grid resistor R_g is large. With small values of R_g, the limiting value that the grid-bias voltage approaches (see Fig. 33) may be near or even above cutoff. Oscillations reach a stable amplitude provided any tendency for oscillations to decrease causes a greater decrease of power consumed than of power sup-

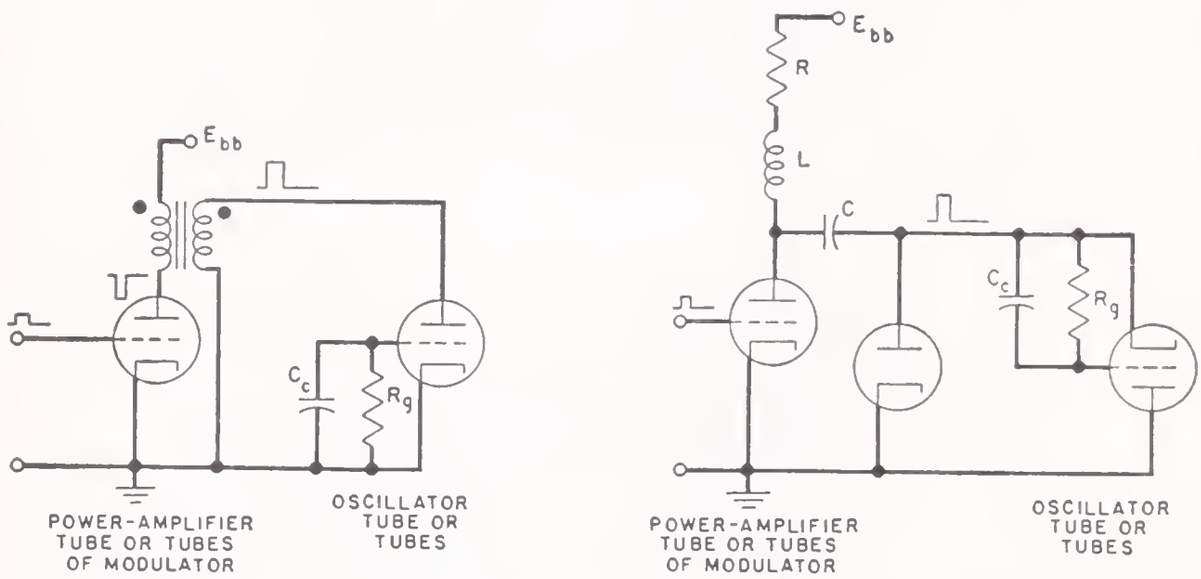


Fig. 34. Coupling of modulator to plate-cathode circuit of triode oscillator (r-f oscillator components omitted).

plied by the plate-current pulses. To ensure self-pulsing, R_g must be large enough so that the limiting grid bias voltage is far below cutoff and C_c must be large enough so that the bias voltage can change only slowly relative to the rate at which the amplitude of oscillation changes.

12. Pulsed Oscillators. One method of pulsing triode oscillators is to apply the plate supply voltage in short pulses. A variety of modulator circuits capable of supplying the required high-voltage, high-current pulses are described in Chap. III. The basic features of circuits for coupling the power-amplifier tubes of a modulator to an oscillator are indicated in Fig. 34. These circuits do not include the r-f-oscillator components and may apply to any oscillator circuit. In Figs. 34*A* and *B*, C_c represents a very small capacitance (the capacitance between a tuning line and the oscillator shielding, or an r-f by-pass capacitor). Therefore C_c permits rapid growth of bias voltage at the beginning of the pulse and does not cause self-pulsing.

The modulator power requirements may be reduced if pulses of voltage are applied to the oscillator grid circuit rather than to the plate circuit. A modulator circuit of this kind employing a cathode follower is indicated in Fig. 35*A*. Plate voltage for the oscillator is always present, but the oscillator tube is maintained cutoff between pulses by the negative voltage E_{cc} of the bias supply. As indicated by the waveform of Fig. 35*B*, oscillations begin when the positive pulse from the cathode follower carries the grid voltage above cutoff. A grid-bias voltage develops rapidly across the small capacitor C_c, and Class C operation is obtained during most of the pulse. The pulse ends when the cathode follower reduces the total

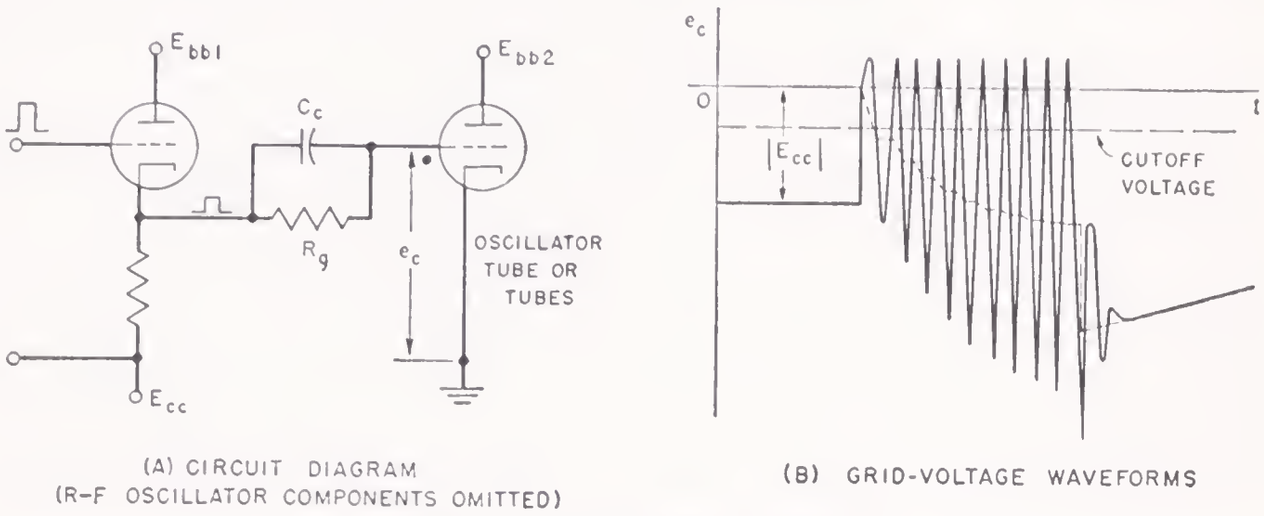


FIG. 35. Grid-pulsed oscillator.

bias voltage far below cutoff, and e_c then returns exponentially to E_{cc} as C_c discharges.

Grid pulsing is not ordinarily employed with oscillator tubes having oxide-coated cathodes, because considerably higher plate supply voltages can safely be used if the plate voltage is applied only during the time of the pulse. With oscillator circuits for which grid pulsing is allowable, however, even greater simplicity than the circuit of Fig. 35*A* can be obtained by the use of self-pulsing. Disadvantages of the self-pulsed circuit include the reduced oscillator efficiency resulting from changing conditions during the pulse and the very critical nature of the adjustments required for proper operation.

A circuit diagram for a self-pulsed radar transmitter (r-f elements omitted) is drawn in Fig. 36. Self-pulsing is obtained, as explained in Art. 11, because of the large values of C_c and R_g employed—about 0.001 to 0.1 μf and 1 to 10 megohms, respectively. A voltage pulse e_k produced by the flow of cathode current in R_k is used to synchronize the remainder of the radar set. Capacitor C_k is small and serves only to by-pass r-f components of the cathode current.

The waveforms of the grid voltage e_c and the synchronizing voltage e_k are indicated in Fig. 37. The waveform of e_c differs from the e_c wave in

Fig. 33, p. 739, only in that a pulsing cycle after steady-state operation has been reached is represented. Because capacitor C_c was charged during the previous pulse, the grid voltage begins at a negative value below cutoff and rises exponentially toward zero. When it reaches a voltage slightly above cutoff, oscillations begin and grow rapidly to the large amplitude indicated at A. Grid current causes the negative bias voltage to increase slowly in magnitude until the value E_1 is reached at point B. At this point, the plate current pulses become too small to sustain oscillations and oscillations cease, as explained in Art. 11.

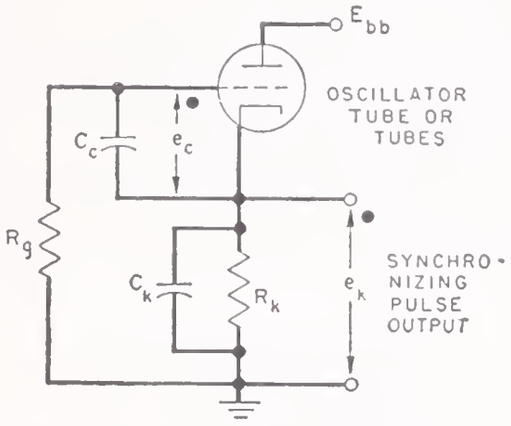


FIG. 36. Self-pulsed oscillator (r-f components omitted).

The e_k waveform of Fig. 37 is determined by the cathode current. The grid voltage rises past the cutoff value shortly before the pulse begins, a very small plate current flows, and therefore e_k has a small value for a relatively long time. The plate current rises suddenly to a high value at the beginning of the pulse and then decreases as the grid-bias voltage builds up. The pulse of e_k is therefore roughly trapezoidal in form. Note that the pulse duration is greatly exaggerated in both waves

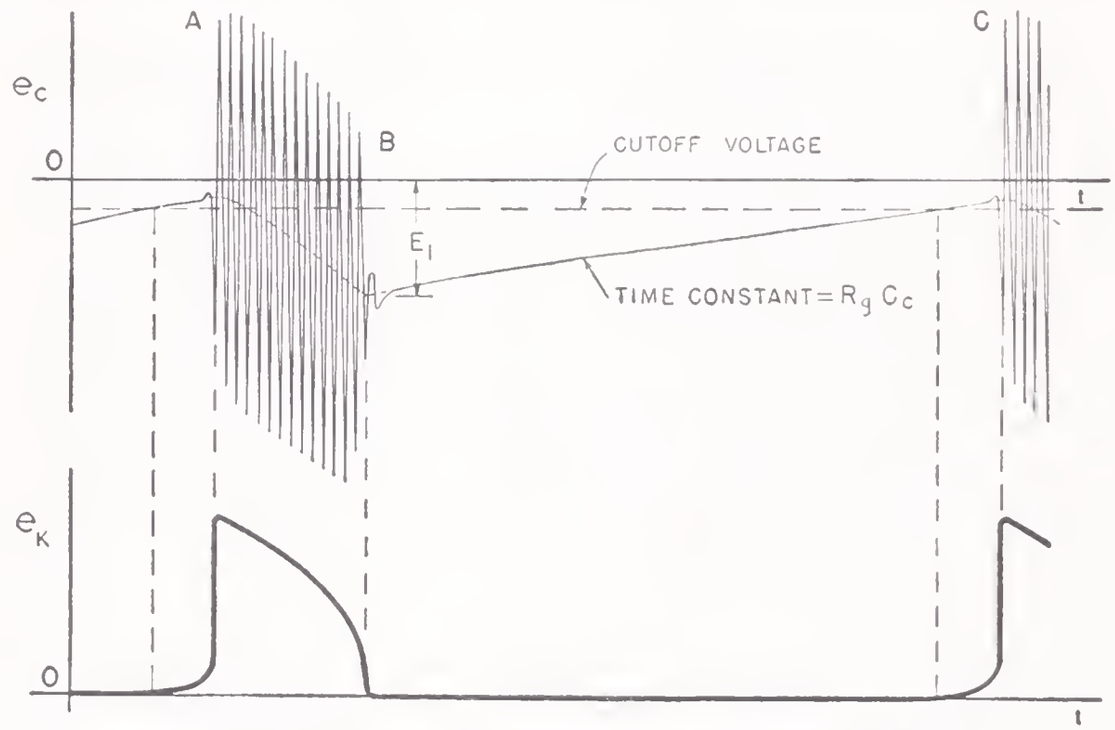


FIG. 37. Waveforms for self-pulsed oscillator of Fig. 36.

of Fig. 37. Typical actual values for radar transmitting oscillators require the pulse to be only about a thousandth of the total period.

Capacitor C_c and the resistor R_k may be adjusted to obtain the desired pulse duration and repetition frequency. The size of the capacitor does not affect appreciably the shape of either the charge or the discharge

curve for the capacitor; instead it determines a scale factor for the time axis for both these curves. Consequently, adjustments of the capacitor affect both the pulse duration and the repetition period, and both of these are changed in nearly the same ratio as the capacitance. Resistor R_g has no appreciable effect upon the pulse duration, since the current through it during the pulse is a very small fraction of the grid current. The grid resistor does, however, control the repetition frequency, because it affects the time constant of the exponential discharge between pulses. Sometimes grid resistors in self-pulsed oscillators are returned to a positive voltage rather than to ground. Changing the point of return affects the repetition frequency but not the pulse duration.

Adjustments of r-f components that affect the amplitude of the oscillations also affect the pulse duration and repetition frequency. For example, if the plate load resistance is decreased, or if the fraction of the r-f plate voltage returned to the grid is decreased, the amplitude of the oscillations obtained at the grid decreases. As a result, the grid current decreases and the pulse length tends to increase, since the smaller grid current takes longer to charge the capacitor. At the same time, however, the voltage E_1, which the capacitor reaches at the end of the pulse, is decreased because the point at which the plate-current pulses supply less power than the load consumes is reached sooner. The decrease of E_1 tends to shorten the pulse and increase the repetition frequency. Therefore, reduction in the amplitude of r-f grid voltage results in a higher repetition frequency and may produce a slightly longer or a slightly shorter pulse. The duty ratio (ratio of pulse duration to repetition period) is increased. A decrease in the plate load resistance causes the average plate current to increase, both because the current during the pulse increases and because the duty ratio increases.

Making adjustments upon a self-pulsed oscillator may cause the circuit suddenly to cease pulsing and to produce a continuous oscillation. Such a change may be caused by an increase of load, a decrease of supply voltage, reduced feedback voltage, or a lower value of R_g, or by any change that makes the oscillation weak so that sufficient bias voltage to end a pulse is not produced. The change to continuous operation causes a tremendous increase of average plate current in radar transmitters because of the large increase of duty ratio; therefore, the fuses or protective relays used with all high-power oscillators are especially needed with self-pulsed circuits.

A number of variations of the circuit of Fig. 36 have been used in radar transmitters. In some circuits a parallel R-C combination in the cathode circuit is used instead of C_c and R_g to build up the bias voltage that causes pulsing. Where the somewhat variable repetition frequency of a free-running self-pulsed oscillator is undesirable, a synchronizing voltage is

introduced into the grid circuit to make definite the pulsing period. Often oscillators make use of a trigger pulse from a modulator tube to start the r-f pulse and then end the period of oscillation because of a bias voltage accumulated on a large grid capacitor. Sometimes an artificial transmission line or pulse network is employed instead of a grid capacitor in order that the oscillator may produce an r-f pulse having a more nearly rectangular envelope.

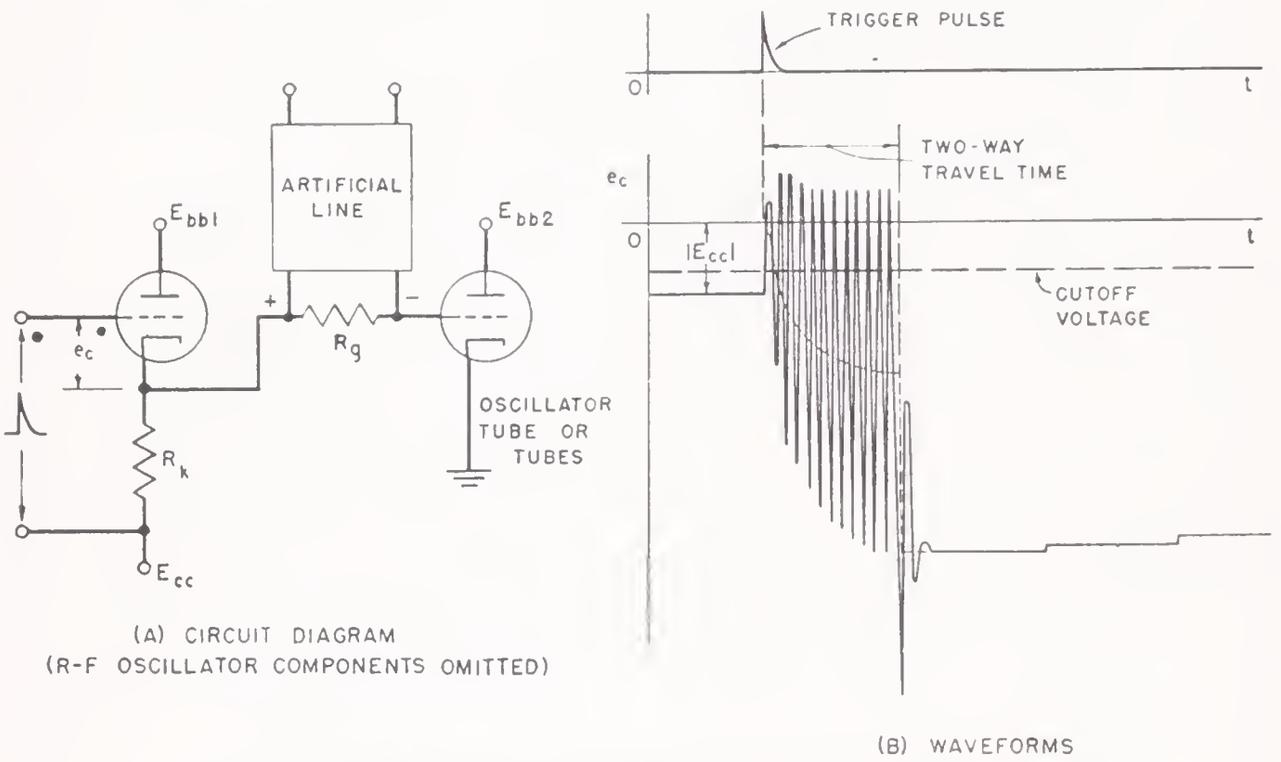


Fig. 38. Oscillator with artificial-line control of pulse duration.

One circuit employing artificial-line control of the pulse is illustrated in Fig. 38*A*. As indicated by the waveforms of Fig. 38*B*, the oscillator tubes are maintained cutoff until the time of the positive trigger pulse from the cathode follower. Oscillations build up within the duration of the trigger pulse, and the resulting grid current initiates a wave on the artificial line, the polarity of the wave being as indicated in Fig. 38*A*. This wave is reflected at the open end of the line, and the reflected wave on reaching the input end of the line causes a negative increment of grid voltage that ends the pulse.

CHAPTER XI

KLYSTRONS AND MAGNETRONS

The microwave counterparts of the triode oscillators of Chap. X are klystron local oscillators and magnetron transmitting tubes. The *reflex klystrons* used in radar are small, moderately low-voltage tubes capable of supplying a few tens of milliwatts of high-frequency power at an efficiency of a few per cent. Radar *magnetrons* are somewhat larger tubes that operate at very high voltages and require the use of rather cumbersome magnets. These magnetrons produce high-frequency oscillations in short pulses of great power, often at an efficiency greater than 50 per cent. Because of the serious limitations of triodes at frequencies in excess of 1,000 Meps, the success of microwave radar is in large measure due to the development of the magnetron and of the klystron-and-crystal form of superheterodyne receiver.

A. KLYSTRONS

The distinguishing feature of klystrons is the special method by which an input voltage controls the movement of electrons in the tubes. Instead of determining the number of electrons that move away from the cathode in each unit of time, the input voltage affects directly only the velocity with which the electrons travel. For this reason, klystrons are called *velocity-modulation* or *velocity-variation* tubes.

1. Two-resonator Klystron.[1] The original type of klystron comprises an electron beam coupled to two cavity resonators. An input signal is applied to one of the resonators, and an amplified version of the signal is obtained from the other resonator. The tube functions as an oscillator if a portion of the output power is returned to the input. Although the two-resonator klystron is not used in radar, it is of interest here as a connecting link between the grid-return triode circuits of Arts. 6 and 7, Chap. X, and the reflex klystrons used as radar local oscillators.

The components of a two-resonator klystron are shown in Fig. 1. The cathode is part of an electron-gun structure for which the accelerator grid, the two resonators, and the collector electrode form the anode. Under the influence of the accelerating voltage E_{acc}, electrons travel along the

[1] W. C. Hahn and G. F. Metcalf, "Velocity-modulated Tubes," *Proc. I.R.E.*, **27** (February, 1939), 106–116; R. H. Varian and S. F. Varian, "A High-frequency Oscillator and Amplifier," *J. App. Phys.*, **10** (May, 1939), 321–327; D. L. Webster, "Cathode-ray Bunching," *J. App. Phys.*, **10** (July, 1939), 501–508.

axis of the tube in a fairly well-defined beam. Some of the electrons are intercepted by each of the five grids, but many reach the collector electrode at the end of the tube. The input and output resonators are of the re-entrant form described in Chap. IX (see Fig. 15, p. 636) and may be thought of as capacitance-loaded coaxial-line resonators made short to obtain high natural frequencies. The flexible diaphragms and a tuning mechanism (not shown in Fig. 1) make possible control of the resonant frequencies by adjustment of the gap spacings.

The input gap is analogous to the cathode-grid space of a triode in that input signals cause an electric field at this gap which brings about a variation or modulation of the electron beam. In a triode, the electron beam is *density modulated;* that is, a dense beam of electrons flows from cathode to plate when the grid voltage is positive, and few electrons (or none at all) traverse the tube when the grid voltage is negative. In the klystron, the electron gun determines the density of the electron beam, and the electric field in the input gap affects only the *velocity* of the electrons that pass through. Electrons that traverse the gap when grid 3 is positive with respect to grid 2 are accelerated by the field, and those which are in the gap a half cycle later are decelerated. Thus the stream of electrons leaving grid 3 is essentially uniform in density but comprises a succession of groups of fast and slow electrons.

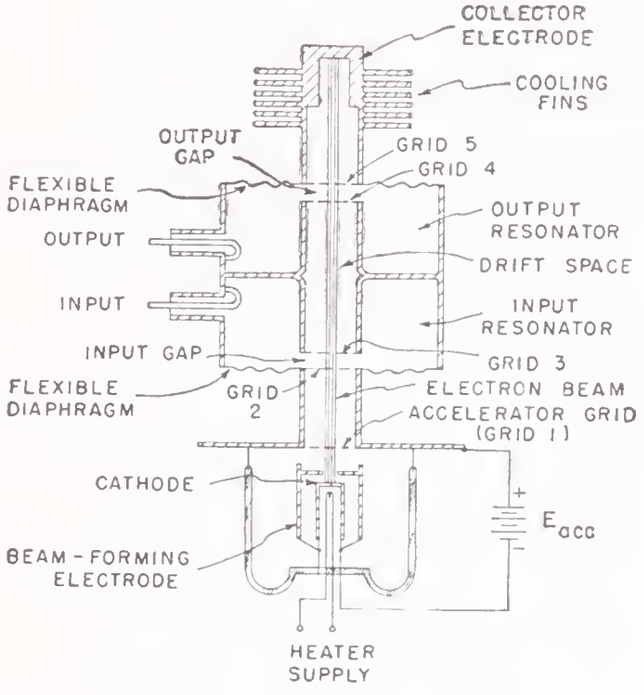
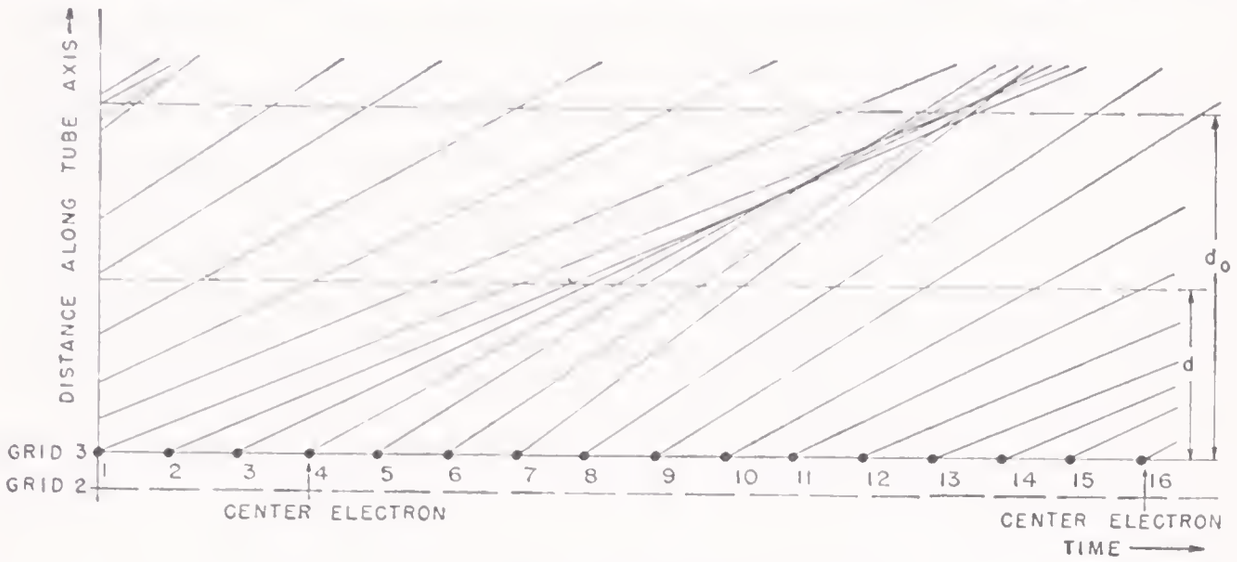


FIG. 1. Cross section of a two-cavity klystron.

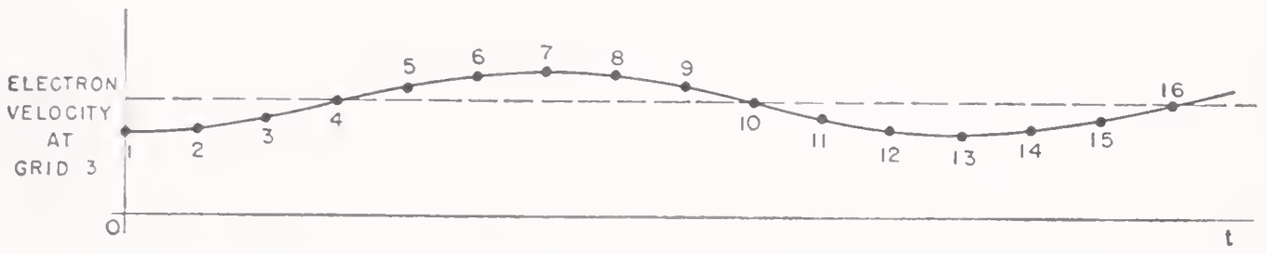
The advantage of the velocity-modulation process is that the electrons are accelerated before they reach the input gap, and therefore very short transit times are possible in gaps of appreciable width. In triodes, on the other hand, electrons leave the cathode with nearly zero velocity and even at the grid attain velocities corresponding to acceleration by only a fraction of the plate voltage. In consequence, the transit time for even the exceedingly small gap (0.004 in.) of the Type 2C40 lighthouse tube is too great for efficient operation in the microwave region.

An apparent disadvantage of the klystron is that the velocity-modulated beam, unlike a density-modulated beam, is not directly usable in the production of output power. This difficulty is overcome by the use of a field-free *drift space* (see Fig. 1) in which the velocity-modulated

beam is converted to a density-modulated beam. Consider the electrons that traverse the input gap as the r-f field is changing from a decelerating field, through zero, to an accelerating field. The decelerated electrons that cross the gap first are overtaken by the electrons that pass the gap when the field is zero, and the accelerated electrons that traverse the gap last overtake the unaccelerated and the decelerated electrons. Thus, in crossing the drift space, the electrons tend to *bunch*, or to arrange them-



(A) APPLEGATE DIAGRAM



(B) VELOCITY-TIME DIAGRAM

FIG. 2. Bunching process in a two-resonator klystron.

selves in groups of increased density centered about the unaccelerated electrons.

The bunching process may be followed in greater detail with the aid of the *Applegate diagram*[1] of Fig. 2. In this diagram the dots 1, 2, 3, and so on, represent instants at which electrons pass grid 3, and the lines are plots of the distance each electron has penetrated into the drift space at any time after leaving grid 3. The lines are straight because the electrons move with constant velocity through the field-free drift space, and the slopes of the lines are proportional to the velocities of the electrons. Thus the steep line for electron 7 in Fig. 2*A* may be correlated with the high velocity for electron 7 in Fig. 2*B*, and the less steep line for electron 1 with the lower velocity of that electron. The convergence of the lines

[1] A. E. HARRISON, "Graphical Methods for Analysis of Velocity-modulation Bunching," *Proc. I.R.E.*, **33** (January, 1945), 20–32.

about the center-electron lines indicates the bunching process, the density variations at any position in the drift space being represented by the distribution of intersections between the slant lines and a horizontal line at the position of interest. Thus, the distribution of intersections along the line at distance d into the drift space indicates appreciable bunching, and optimum bunching occurs at the distance d_o.

Because of the bunching in the drift space, the flow of electrons through the output gap is analogous to the flow of plate current through the grid-plate space of a triode. In fact, if optimum bunching occurs at the output-gap position, the klystron may be compared with a Class C amplifier —the electrons move through the output gap in short-duration pulses, only a few electrons traversing the gap in periods between pulses. In a triode, both a-c and d-c field components are present in the grid-plate space, and the electrons simultaneously receive energy from the d-c field and deliver it to the a-c field. In a klystron the two fields are separated. Electrons receive energy from the d-c source in the accelerated motion between cathode and grid 1. Part of this energy is delivered to the output-cavity oscillations when electrons are decelerated by the fields in the output gap. The remainder of the kinetic energy of the electrons is converted to heat when the electrons strike the collector electrode.

Energy interchanges between electrons and resonator fields occur also at the input gap. Electrons accelerated in this gap receive energy, and those decelerated deliver energy to the field. To the extent that electrons at the input gap (in contrast to those at the output gap) move in a stream of uniform density, as many electrons are accelerated as are decelerated, and the net energy gained or lost by the input resonator is zero. If the transit time for the input gap is appreciable, some density modulation occurs within the gap, and an energy transfer from resonator to electrons results (which may be compared with grid loading in a triode). Loading of the input circuit is one reason for maintaining a small transit time in the input gap.

Another reason for maintaining small transit time, which applies to both input and output gaps of klystrons, is that the coupling between the electron beam and the resonators is decreased if the transit time is appreciable. An appreciable input-gap transit time reduces the amount of velocity modulation produced by a given r-f voltage at the gap, because no electron can receive maximum acceleration or deceleration during more than a fraction of the time it is in the gap. Similarly, appreciable output-gap transit time reduces the power delivered to the resonator by a given beam current, because an electron can be acted on by the maximum decelerating field during only a portion of the output-gap transit time. The net effect is as if step-down ideal transformers were interposed between the resonators and the electron beam.

Despite these effects the transit time for klystron gaps is usually of the order of half the period of the r-f oscillation. In the choice of gap spacings a compromise is made between the long transit times obtained with wide spacings and the large intergrid capacitances that result from narrow spacings. A large capacitance reduces the impedance-level parameter $\sqrt{L/C}$ for a resonator and therefore causes large circuit losses for a given r-f voltage at the gap (see Arts. 6 and 13, Chap. IX).

From Fig. 2*A* it may be seen that the extent of the bunching obtained at any fixed position assumed for the output gap is a function of the r-f voltage at the input gap. Thus the lines of Fig. 2*A* are parallel and no bunching occurs if the input voltage is zero. Increasing the input voltage increases the differences of slope among the lines and thus increases the r-f current at the output gap caused by density modulation. For small input voltages, the fundamental component of the output current is very nearly proportional to the input-voltage amplitude, and the output-current to input-voltage ratio may be represented by a *transconductance* value analogous to the transconductance of ordinary amplifier tubes. If the input-voltage amplitude is increased, the proportionality eventually fails; in fact the output current is limited to the value corresponding to optimum bunching at the output gap and is reduced for input voltages greater than that for optimum bunching.

The transconductance of a klystron differs from that of low-frequency tubes in that it can be varied over wide ranges through control of the accelerating voltage. Low accelerating voltages and long drift spaces produce appreciable bunching with small input voltages and thus provide high transconductances. The attainable transconductance is limited, however, because relatively high accelerating voltages are required for short gap transit times, and too long a drift space permits *debunching* (spreading of the electron bunches because of the mutual repulsion of the negatively charged electrons) to occur. Despite these limitations power amplifications by factors in excess of 100 are obtainable from klystrons at frequencies in the microwave region.

The two-resonator klystron, nevertheless, is *not useful as an r-f amplifier* in microwave radar systems. The reason is that the purpose of radar r-f amplifiers is to decrease the noise figure of a receiver by amplifying the signal before the noise of the mixer and first i-f amplifier is introduced (see Art. 17, Chap. VI). The klystron, unfortunately, is a relatively powerful noise generator and would increase rather than decrease the noise figure of the receiver. Two effects contribute to the high noise generation in klystrons relative to low-frequency amplifiers. Noise in low-frequency amplifiers is the result of density modulation of the electron stream emitted from the cathode, and this density modulation is considerably reduced by the smoothing effect of space charge (Art. 16,

Chap. VI). One reason for the increased noise in klystrons is that the smoothing effect of space charge is considerably reduced because of the high frequency of operation and because of the interception effects of the several klystron grids. The other reason is that the thermal velocities with which electrons leave the cathode constitute a velocity modulation of the electron beam which is converted to density modulation during the traversal of the cathode-to-input-gap region, as well as in the drift space.

The klystron of Fig. 1 can be made to oscillate if the output and input coupling loops are joined by a coaxial line. The device thus obtained is very closely analogous to the grid-return triode oscillator of Fig. 18, p. 722. Because of the feedback line, oscillations of the output cavity cause an r-f voltage at the input gap, and this voltage results in bunches of electrons traversing the output gap. These electrons contribute energy to sustain the oscillations, provided that the phase of the feedback is such that the majority of the electrons pass the output gap at the time of a decelerating field. A disadvantage of the two-resonator oscillator is the difficulty of adjustment. In order that sufficient voltage may be developed at the input gap to provide appreciable feedback, the input resonator must be tuned to very nearly the same frequency as the output resonator. Oscillations are then obtained only if the phase shift in the feedback line and the phase delay in the drift space are adjusted to provide the correct feedback phase. The adjustment for oscillation is somewhat simplified in the *floating-drift-tube klystron*, which employs (instead of the two resonators and feedback line) a single resonator having two points of entry, one at each of the klystron gaps. An even simpler oscillator, and one especially suited for use as a local oscillator, is the *reflex klystron* discussed in the next article.

2. Reflex Klystrons. The reflex klystron[1] differs from the tube of Fig. 1 in that the output cavity is omitted and, instead of the collector electrode, a *reflector electrode* is placed a short distance from the single cavity that remains. The reflector is maintained at a negative voltage of several hundred volts with respect to the cathode. Electrons that pass through grids 2 and 3 (see Fig. 1) therefore encounter a strong decelerating field. The electrons are slowed to a stop, accelerated in the reverse direction, and returned through the cavity grids toward the cathode. Because of a bunching process that occurs in the reflecting field, density modulation is present during the return passage through the cavity gap, and the single cavity performs the functions of both the input and the out-

[1] J. R. PIERCE, "Reflex Oscillators," *Proc. I.R.E.*, **33** (February, 1945), 112–118; E. L. GINZTON and A. E. HARRISON, "Reflex Klystron Oscillators," *Proc. I.R.E.*, **34** (March, 1946), 97–113; D. R. HAMILTON, J. K. KNIPP, and J. B. H. KUPER, *Klystrons and Microwave Triodes*, 311–526 (McGraw-Hill Book Company, Inc., New York, 1948).

put cavities of the two-cavity tube. If the r-f energy extracted from the electrons on the return trip exceeds that lost in the cavity and in the velocity-modulation process, the excess energy is available as useful output power and the tube functions as an oscillator.

The bunching process that occurs in the field between cavity and reflector is slightly different from the corresponding process in a field-free drift space. Electrons that are accelerated most in the initial passage between the cavity grids penetrate farthest into the reflection field and take longest to return to the cavity. Thus a bunch is formed about an unaccelerated electron that traverses the gap at an instant when the r-f field is changing from the accelerating to the decelerating direction. The slower electrons that leave the cavity after the instant of zero field intensity overtake the faster electrons that leave at an earlier instant. The electron motions may be likened to the motions of balls thrown upward against the gravitational field of the earth. A ball thrown with high velocity travels farther than a ball thrown with lower velocity and requires a longer time for the ascent and descent. Thus the two balls may be made to return together if the throwing of the low-velocity ball is delayed an appropriate time interval after that of the high-velocity ball.

The reflex-klystron bunching process may be explained in more detail by means of an Applegate diagram that takes into account the effect of the reflecting field. Such a diagram is drawn in Fig. 3. The velocities of the electrons leaving the cavity are indicated in Fig. 3*B*, and the distance traversed by each electron is plotted as a function of time in Fig. 3*A*. If the reflecting field is uniform, the electrons experience a constant acceleration toward the cavity and the curves of Fig. 3*A* are parabolas. Electron 4 leaves the cavity as the r-f field changes from the accelerating to the decelerating direction, and thus a bunch forms about this electron. Electron 3 leaves the cavity with higher velocity, penetrates farther into the retarding field, and returns to the cavity at essentially the same instant as electron 4. Similarly, electron 5 leaves with lower velocity, penetrates a shorter distance, and returns at essentially the same instant as electrons 3 and 4.

The energy delivered by the returning bunch to the r-f oscillations of the cavity, and consequently the possibility of the tube functioning as an oscillator, depends upon the phase of the r-f oscillations at the instant that the bunch returns to the cavity. For maximum delivery of power, the field encountered by the returning electrons should be the peak value of the r-f field in the direction to decelerate the electrons. To produce this condition, the unaccelerated electrons (about which bunches form) should remain in the reflecting field for three-quarters of a cycle of the r-f oscillation or any integral number of cycles plus three-quarters. For example, in Fig. 3*A*, the time in the retarding field is three-quarters of a

cycle, and the bunch returns when the cavity field has a maximum value in the direction to accelerate electrons *away* from the cathode. Such a field decelerates the returning bunch and gains energy from the electrons.

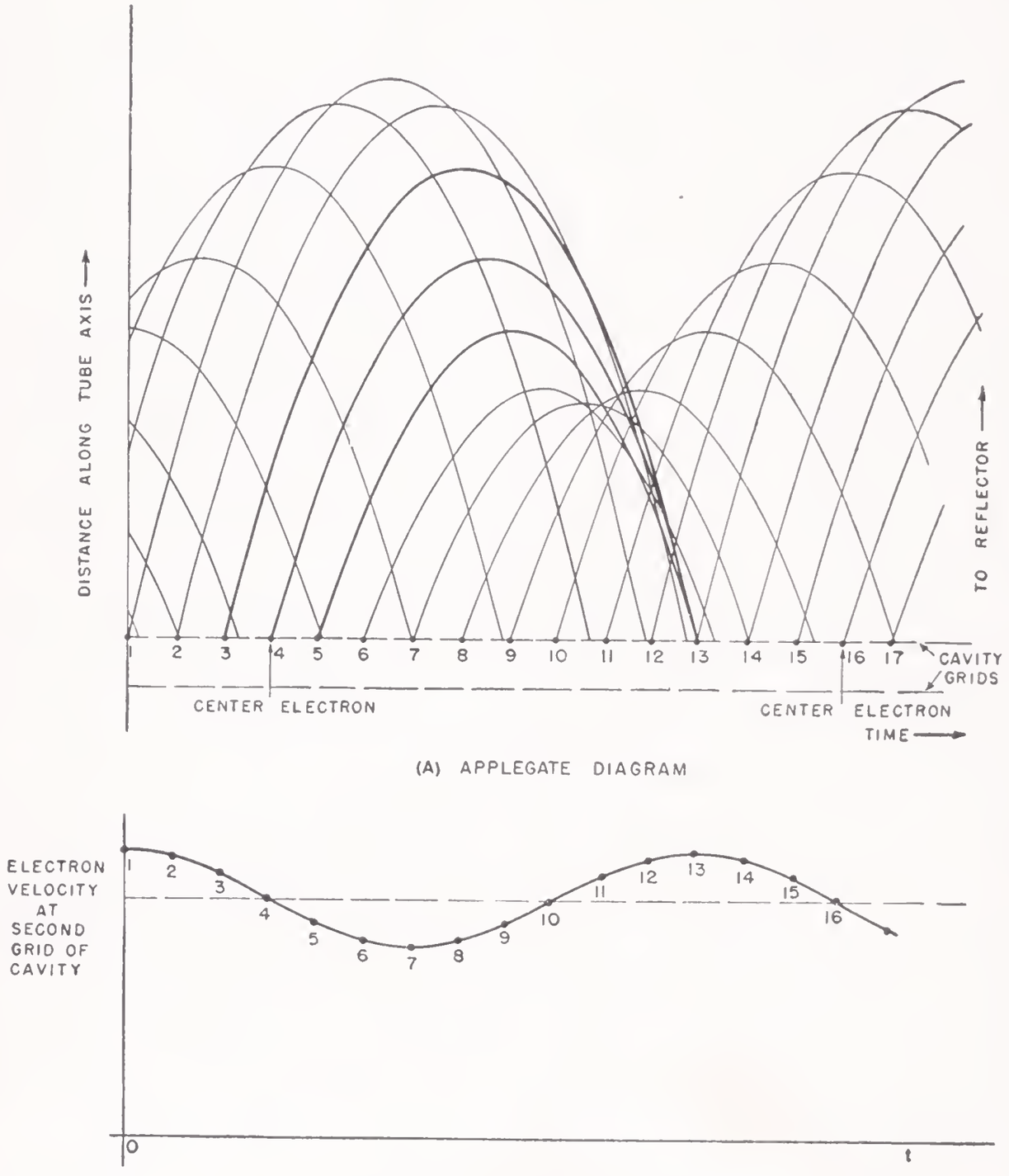

DISTANCE ALONG TUBE AXIS
TO REFLECTOR
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
CAVITY GRIDS
CENTER ELECTRON
CENTER ELECTRON
TIME
(A) APPLEGATE DIAGRAM
ELECTRON VELOCITY AT SECOND GRID OF CAVITY
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
0
t
(B) VELOCITY-TIME DIAGRAM

FIG. 3. Bunching process in reflex klystron.

The time spent by the electrons in the reflecting field can be controlled by adjustment of the reflector voltage. For a negative reflector voltage of large magnitude, the field is strong and the penetration distance short. Thus the three-quarter-cycle delay time corresponds to a large-magnitude negative reflector voltage, and the $1\frac{3}{4}$, $2\frac{3}{4}$, $3\frac{3}{4}$, . . . , times correspond to a series of reflector-voltage values of decreasing magnitude. Oscilla-

tions are thus obtained for a series of different reflector-voltage adjustments, and these different manners of oscillation are known as *voltage modes* of the klystron. The voltage modes should not be confused with resonant modes of the klystron cavity. Only one cavity mode is

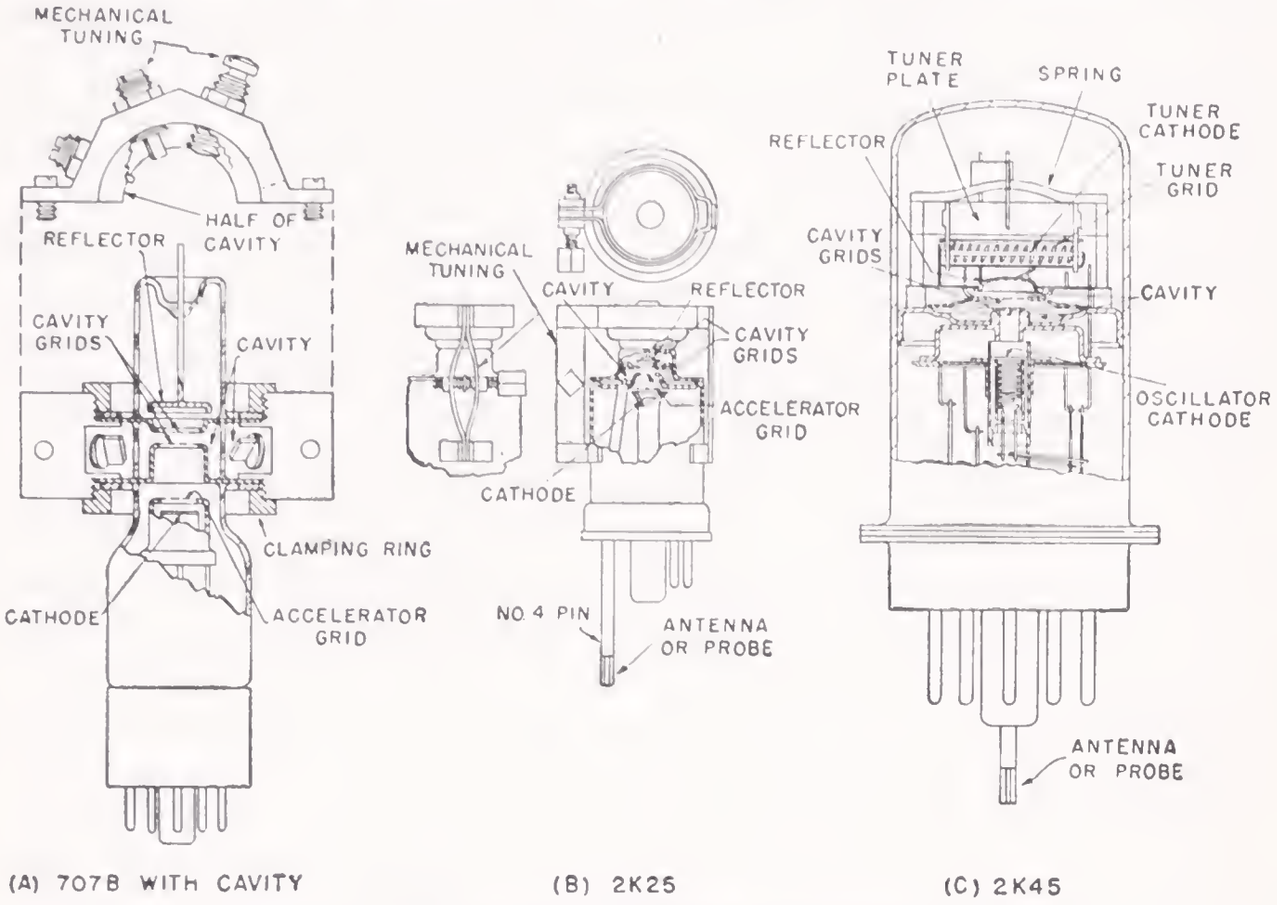


FIG. 4. Construction of typical reflex klystrons.

employed, and the center frequency of all voltage modes is the same, corresponding to the natural frequency of the cavity mode.

After passing back through the cavity, most of the electrons have insufficient energy remaining to return to the cathode. The electrons there-

TABLE 1. REFLEX KLYSTRONS

Type No.	*Frequency range, Mcps*	*Accelerating voltage, volts*	*Beam current, ma*	*Reflector voltage, volts*	*Power output, mw*	*Electronic tuning range, per cent*
2K28/707B	*	300	25	−175	150	0.7
726A	2800–3100	300	22	−160	100	0.7
417A	2700–3300	1000	40	−250	400	0.2
2K25, 723A/B	8500–9600	300	22	−160	25	0.5
2K45	8500–9600†	300	22	−160	30	0.5
2K50	23,300–24,700†	300	15	−100	15	0.3

* Depends upon external cavity. Typical cavity can be adjusted by tuning screws or paddles over range of 2,500 to 3,800 Mcps.

† Thermally tuned integral cavity.

fore either pass to the sides of the tube and return through the accelerating-voltage source or pass on to the control grid and back through its biasing source to the cathode.

The construction and characteristics of typical reflex klystrons are indicated by the drawings of Fig. 4 and the data of Table 1. The table indicates, for each tube type, the ranges over which the oscillation frequency can be adjusted by change of the cavity dimensions and a set of typical operating conditions. Included in the operating conditions is the electronic tuning range which refers to the frequency shift that may be obtained by adjustment of the reflector voltage within a small range about the optimum value for the voltage mode in use. The power output decreases as the reflector voltage is altered from the optimum value, and the tuning range is measured between the half-power frequencies. The mechanism of electronic tuning is discussed in the next article.

The Type 707B klystron of Fig. 4*A* is an example of the use of an external cavity with a glass-envelope tube employing disk seals for the cavity grids. The external cavity makes possible the use of the tube over a very wide frequency range but is in general rather bulky and inconvenient in radar-local-oscillator service. The Type 2K28 tube of Table 1 has electrical characteristics essentially identical with those of the 707B, but the glass envelope is somewhat shorter.

The Type 2K25 klystron of Fig. 4*B* (see also Fig. 24, p. 38, in Chap. I) is identical, except for more rigid test specifications, with the Type 723A/B and is very widely used as the local-oscillator tube in 3-cm radar systems. The integral cavity is tuned mechanically by the adjustment of bowed struts. This adjustment tilts the top of the tube, flexes a portion of the upper wall of the cavity, and changes the spacing of the cavity grids. Output power is obtained from a small-diameter coaxial line occupying the position of one of the pins of the octal socket. The line is loop-coupled to the cavity and is terminated in a small antenna or probe. Ordinarily the tube is mounted so that the antenna projects into a 3-cm waveguide, thus forming a probe type of coaxial-line-to-waveguide junction (see Art. 17, Chap. IX). The Type 726A tube of Table 1 is almost identical in external appearance with the 2K25 of Fig. 4*B*. The cavity height is larger in the 726A, so that oscillations in the wavelength region of 10 cm are obtained, and the output line terminates in a fitting for connection to an additional length of coaxial line.

The Type 2K45 tube is similar in electrical characteristics to the 2K25 but, as shown by Fig. 4*C*, is of rather different construction because of the incorporation of a thermal tuning mechanism to permit remote control of frequency. The upper portion of the tube is a triode in which the length of the plate changes when the temperature of the plate is altered by a change in plate current. The length of the plate controls the amount of

bow in the spring mounted above the plate, and vertical motion of the center of the spring is communicated by a yoke and two rods to the upper wall of the cavity. Thus the spacing of the cavity grids can be controlled through control of the triode grid voltage. The thermal tuning can operate only slowly because of the time required for the plate to heat and cool, but in combination with rapid tuning over a small range by reflector-voltage adjustment, it makes possible complete control of the frequency of oscillation.

The Type 2K50 klystron of Table 1 operates in the 1-cm wavelength region and employs parts having very small and extremely accurate

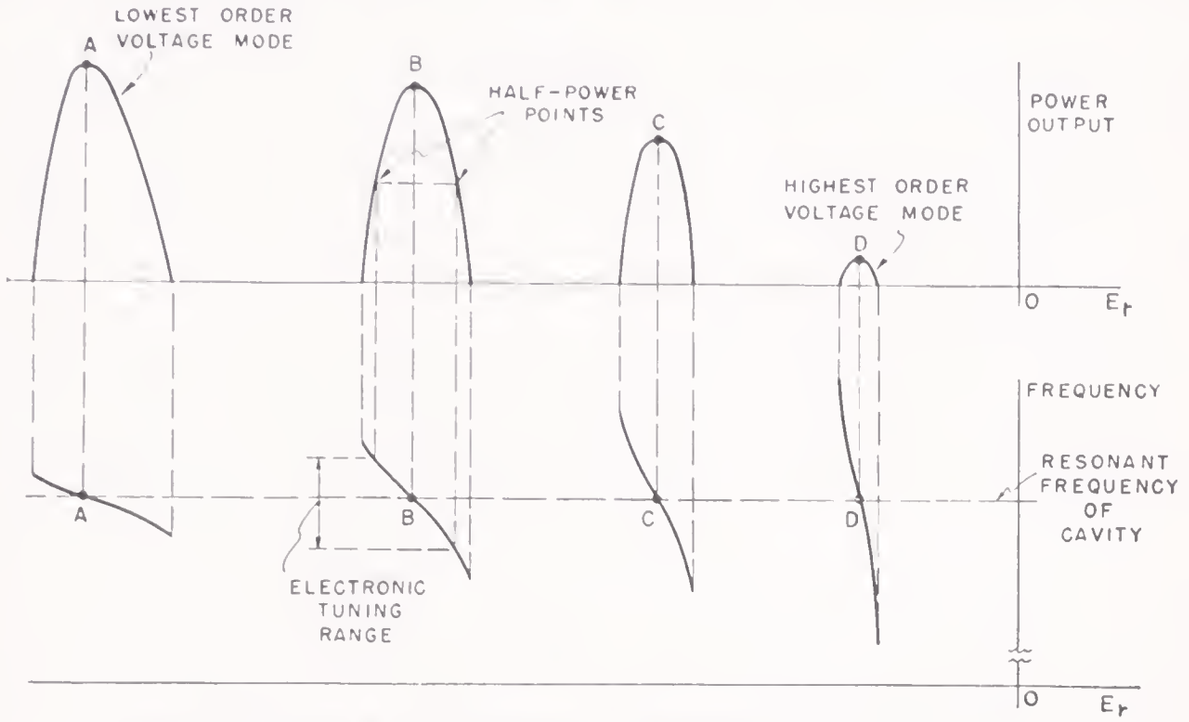


FIG. 5. Power and frequency characteristics of reflex klystron.

dimensions. Output is obtained through an aperture in the cavity and a tapered section of waveguide. Like the 2K45, the 2K50 includes a triode section for thermal tuning. The 417A klystron is included in Table 1 for its general interest although, because of its relatively great power output and its small electronic tuning range, it is more useful as a low-power transmitter than as a local-oscillator tube. When operated with reduced accelerating voltage and beam current, this tube can produce sufficient power for local-oscillator service, but its size and small electronic tuning range make it less convenient to use than other tubes in Table 1.

3. Characteristics of Reflex Klystrons. The variations of power output and frequency that occur as the reflector voltage E_r of a reflex klystron is changed are particularly important characteristics of the tube and may be represented by curves such as those of Fig. 5. These curves are somewhat idealized versions of curves that may be obtained experimentally if the power and frequency of the oscillations are measured as the reflector voltage is varied from zero to a large negative value, acceler-

ating voltage and load remaining fixed. Oscillations occur only for certain ranges of the reflector voltage (corresponding to the voltage modes) for which the electron bunches return to the cavity in the proper phase to deliver energy to the cavity.

The center points of the modes, labeled A, B, C, and D in Fig. 5, correspond to reflector voltages for which the time spent by electrons in the retarding field is exactly an integral number of cycles plus three-quarters of a cycle. At these points the oscillation frequency is the resonant frequency of the cavity, and the power output is the maximum power of the mode. Observe the general tendency for the power output to be less and the frequency range to be greater for the modes of higher order. (The order of a mode is the integer to which three-quarters must be added to obtain the number of cycles spent by electrons in the reflecting field. Thus higher order modes occur at lower magnitude reflector voltages and correspond to longer times available for electrons to form bunches.)

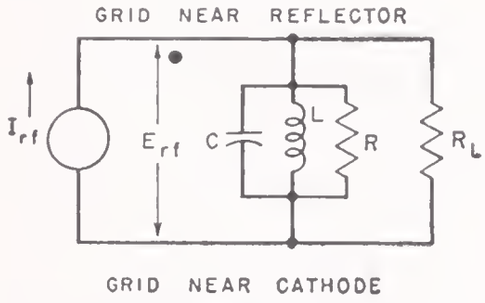


FIG. 6. Equivalent circuit for reflex klystron.

The reasons for the shapes of the curves in Fig. 5 may be explained by reference to the equivalent circuit of Fig. 6. The resonant cavity of the klystron is represented by a parallel L-C circuit, and a resistor R is added to account for the cavity losses, including the loading effect of the electron beam. The load coupled to the cavity is assumed, in this article, to be equivalent to a shunt resistor R_L, and the current of the bunched electron beam returning through the cavity is represented by the current source I_{rf}. In accordance with the discussion of Art. 13, Chap. IX, the current source is connected in parallel with the equivalent circuit of the resonator, referred to terminals at the cavity grids. Note that the reference directions in Fig. 6 are taken so that positive values of the r-f cavity voltage represented by E_{rf} correspond to an electric field that accelerates electrons away from the cathode, and positive values of the current represented by I_{rf} correspond to positive current directed away from the cathode or to an electron bunch returning toward the cathode. Under normal operating conditions, the waveform of the electron-beam current is by no means sinusoidal but comprises a series of narrow pulses. The fundamental component of the wave is represented by the complex number I_{rf}; the other Fourier components produce inappreciable voltage between the cavity grids and may be ignored.

If the voltage E_{rf} is zero, all electrons follow the same paths, no bunches are formed, and I_{rf} also is zero. If E_{rf} has some small value because of noise or any transient disturbance, a slight tendency for the electrons to bunch occurs and the current I_{rf} has a correspondingly small value. If

the reflector voltage is adjusted for operation at the center of a mode, I_{rf} is in phase with E_{rf} and the power $|I_{rf}|\,|E_{rf}|$ is supplied to the cavity. Provided the power loss to R and R_L is less than this amount, the r-f oscillations grow in amplitude because of the excess power supplied, and the larger value of $|E_{rf}|$ causes an increased tendency for the electrons to bunch and a larger value of $|I_{rf}|$. Thus the power supplied, as well as the power lost, increases as the oscillation grows. As long as $|I_{rf}|$ increases in proportion to $|E_{rf}|$, the excess of input power remains and the oscillation amplitude continues to increase.

The magnitude of E_{rf} may be considered to control the position along the tube axis at which the electron beam is bunched to the greatest possible extent. For small values of $|E_{rf}|$ the bunches are not completely formed until after the returning beam has passed through the cavity, and the underbunched beam in the cavity yields a small value of $|I_{rf}|$. A larger value of $|E_{rf}|$ corresponds to *optimum bunching* at the position of the cavity and results in the largest possible value of $|I_{rf}|$. Still larger values of $|E_{rf}|$ cause the bunches to form too soon and to spread out somewhat at the cavity position. The beam in the cavity is then *overbunched*, and a decreased value of $|I_{rf}|$ results. This failure of $|I_{rf}|$ to increase indefinitely with increase of $|E_{rf}|$ limits the amplitude reached by the growing oscillations. Ordinarily some overbunching occurs in the final steady-state condition, the exact degree of bunching being dependent upon the adjustment of R_L.

If the tube is operated in a high-order mode, the electrons penetrate relatively far into the reflecting field and remain in the field a relatively long time. Thus a given degree of bunching is obtained with relatively low r-f cavity voltage. For operation in a low-order mode, a shorter bunching time is available and a larger r-f voltage is required to provide the same degree of bunching and the same value of $|I_{rf}|$. The change in mode order is thus analogous to a change in the amount of feedback in a triode oscillator. The r-f cavity voltage corresponds to the plate voltage developed in a triode oscillator, and the current I_{rf} produced by the bunching process corresponds to the r-f plate current produced because of the feedback connection. An increase in the mode order is equivalent to an increase in feedback because higher values of $|I_{rf}|$ are produced by a given voltage E_{rf}. Correspondingly, it may be expected that oscillation may be impossible because of inadequate feedback in the lowest order modes, especially for low values of load resistance, and that the power output obtainable in high-order modes may be small because of the excess feedback present.

If the same degree of bunching is to be obtained at the center point of each mode, different values of load resistance must be employed for the different modes, lower values of R_L being required in the high-order modes

than in the low-order modes. If the load resistance is adjusted for a fixed degree of bunching, the power delivered by the electron beam decreases steadily with increase of mode order. The power output is less than the power delivered by the electrons because of the loss in R, and this loss is most important in the modes of lowest order. Thus the power output may increase from the lowest order mode obtainable to the next lowest mode, but ultimately the power decreases with increase of mode order.

The curves of Fig. 5 apply for a fixed adjustment of load resistance. Therefore, in the high-order modes R_L is greater than the optimum value, a highly overbunched condition is obtained, and the power output is reduced because not only $|E_{rf}|$ but also $|I_{rf}|$ is small. In the low-order modes the power output is also reduced because R_L is smaller than the optimum value of load resistance. Usually several of the low-order modes can be made to appear or disappear in succession by adjustment of the load resistance.

The curves of Fig. 5 indicate that a shift of the reflector voltage in either direction from the center point of a mode causes a reduction of the power output and a shift of frequency away from the resonant frequency of the cavity. These effects are the same as those obtained in triode oscillators if the phase of the feedback voltage is shifted so that the pulses of plate current do not occur at the instant of minimum plate voltage. An increase in the magnitude of the negative reflector voltage causes the electron bunches to return at an earlier instant and corresponds to a shift of the feedback phase in the leading direction; a decrease in the reflector-voltage magnitude delays the return of the bunches and corresponds to a lagging feedback phase.

Both the klystron and the triode oscillators (see Art. 2, Chap. X) may be thought of in terms of a swinging pendulum sustained in motion by the application of a sharp pulse of force once each cycle. If the pulse is applied at the instant of maximum velocity (as the pendulum passes through its central position), the oscillations are sustained with no effect upon frequency. If the pulse is applied at an earlier instant, it is slightly less effective in supplying power to sustain the oscillations, and at the same time it speeds the motion of the pendulum toward its central position and thus increases the frequency of oscillation. A pulse applied after the pendulum has passed its central position is also less effective in supplying power and decreases the frequency of oscillation by extending the swing that occurs after the pulse is applied. In like manner, the power output of a klystron (or triode) is reduced if the electron bunches move between the cavity grids (or through the grid-plate space) at an instant when the decelerating field is not at its maximum value. The frequency is increased if the electrons traverse the field before the instant

of maximum field intensity or reduced if the traversal occurs after the maximum-field instant.

These effects may be explained more directly for the klystron in terms of Fig. 6 and the phase angle θ between the current I_{rf} and the voltage E_{rf}. The current may be resolved into components $|I_{rf}| \cos \theta$ and $|I_{rf}| \sin \theta$, respectively in phase and in quadrature with E_{rf}. The in-phase component delivers power to the oscillations and must equal the total current supplied to R and R_L. The quadrature component is in the same phase as current in L or C, and the circuit must operate at such a frequency that the total current supplied to L and C in parallel equals the quadrature component. At the center point of the mode, the quadrature component is zero, the total current to L and C must be zero, and resonant-frequency operation occurs. For a larger magnitude negative reflector voltage, I_{rf} leads E_{rf} and the quadrature component has a leading phase. Thus the leading current supplied to C must exceed the lagging current in L, and the frequency of operation is increased. At the same time, the in-phase current $|I_{rf}| \cos \theta$ is decreased, so that the power output is less.

The decrease of voltage $|E_{rf}|$ obtained as the power output is reduced by shift of the reflector voltage from the value at the mode center alters the bunching conditions. In a high-order mode, the electron beam is strongly overbunched at the mode center, so that a tendency for $|I_{rf}|$ to increase with decrease of $|E_{rf}|$ exists, and a large phase angle between I_{rf} and E_{rf} can be introduced by reflector-voltage change before appreciable loss of power occurs. Thus large values of the quadrature current component $|I_{rf}| \sin \theta$ may be obtained, corresponding to large frequency shifts or a wide electronic tuning range. In a low-order mode, on the other hand, appreciable overbunching is not present even at the center of the mode, and a small phase angle between I_{rf} and E_{rf} reduces the power output greatly. Thus the electronic tuning range decreases with decrease in mode order.

The effect of changes in $|E_{rf}|$ upon the formation of bunches is associated also with the asymmetry of the low-order modes noticeable in Fig. 5. If the reflector-voltage magnitude is increased, the shorter time available in the retarding field adds to the effect of reduced $|E_{rf}|$ in decreasing the tendency of the electrons to bunch. If the reflector-voltage magnitude is decreased, the longer time in the retarding field partly offsets the effect of reduced $|E_{rf}|$. Thus the power output falls less rapidly for the decrease of reflector-voltage magnitude, and each mode is wider on the low-frequency side of the center point than on the high-frequency side. The asymmetry is noticeable, however, only in the low-order modes because the percentage change in bunching time that can occur within a high-order mode is negligible.

4. Electronic Tuning, Hysteresis, and Load Effects. For use as the local-oscillator tube in a radar receiver, a klystron need not supply large amounts of power but should oscillate at a frequency that is stable and easily controlled. Output power in excess of 20 mw is available from most reflex klystrons (see Table 1, p. 753), and this power is sufficiently greater than the power required by a mixer crystal to permit use of loose local-oscillator coupling and therefore to prevent loss of signal power to the oscillator (see Art. 28, Chap. VI). Frequency stability depends upon the reduction of frequency shifts caused by supply-voltage changes, by load changes, by ambient-temperature changes, and (in air-borne equipment) by atmospheric-pressure changes. Regulated power supplies are customarily employed, and the loose coupling of the oscillator is of assistance in making the frequency independent of mixer-circuit impedances. Considerable effort is expended in klystron design to provide cavities and tuning mechanisms that are as nearly as possible unaffected by temperature and pressure changes. Despite these precautions, changes of klystron frequency do occur during operation; and to counteract these changes as well as changes of transmitter frequency, automatic-frequency-control circuits are employed (see Art. 30, Chap. VI). In order that the afc circuits may function properly, as well as to permit convenient manual frequency control, the frequency of a klystron local oscillator should be smoothly adjustable over a wide range by means of reflector-voltage changes.

The need for a wide electronic tuning range suggests the use of a voltage mode of high order. However, if a mode of excessively high order is selected, the power available is too small even for local-oscillator applications, and a compromise between wide range and power is necessary. Use of a very high-order mode is undesirable also because the noise output of a klystron is essentially the same for all voltage modes. Thus the closer coupling to the mixer required with high-order, low-power modes increases the noise at the crystal and increases the receiver noise figure (see Art. 17, Chap. VI). Ordinarily a mode of order 2 to 4 (corresponding to a time of $2\frac{3}{4}$ to $4\frac{3}{4}$ cycles in the reflecting field) is found suitable. The width of the tuning range depends not only upon the mode employed but also upon the design of the klystron cavity. A wide spacing between cavity grids results in a very small capacitance C in the equivalent circuit of Fig. 6 and a correspondingly large value of L. For such a circuit, the frequency shift provided by the quadrature current $|I_{rf}| \sin \theta$ is large and wide tuning ranges are possible. Here, again, however, a compromise is necessary, because excessively wide gaps result in long transit times and make the quadrature current less effective in producing a frequency shift. Electronic tuning ranges (between half-power frequencies) of about 0.5

per cent of the center frequency are usual in local-oscillator klystrons (see Table 1, p. 753).

Of interest in addition to the width of the tuning range is the rate at which tuning occurs, or the slopes of the frequency curves in Fig. 5. This tuning rate depends upon the mode used, varies from one point to another in a mode, and (as is to be explained) is also a function of the load impedance employed. In 10-cm tubes, the tuning rate is usually about 0.5 Mcps per volt, and for 3-cm tubes, 2.0 Mcps per volt is a typical figure. Although this dependence of frequency upon voltage is desirable because it makes electronic tuning possible, it is undesirable from the point of view of power-supply requirements. Ordinarily, the cavity is placed at ground potential, and negative voltages are supplied to both the cathode and the reflector. Frequency shifts are caused by cathode-voltage changes as well as by reflector-voltage changes, the effect of cathode-voltage changes being about half as great as that of reflector-voltage changes and in the opposite direction. When automatic-frequency-control circuits are employed, it is possible to omit power-supply regulation and to depend upon the afc circuit for correction of the resulting frequency shifts. This practice, however, reduces the range of electronic tuning available for correction of frequency shifts arising from other causes, and close regulation of both cathode and reflector supply voltages is therefore desirable.

For proper operation of frequency-control circuits, a smooth, uniform variation of frequency with reflector voltage is desired. Unfortunately, unless special precautions are taken in the design of the tube and in the choice of a load, power and frequency curves much more irregular than those of Fig. 5 may be obtained. In particular, an effect known as *hysteresis* may be observed—the power and frequency curves measured as the reflector voltage is *increased* through the mode may differ from corresponding curves measured with a *decreasing* reflector voltage.

One cause of hysteresis and mode asymmetry is a dependence of the phase of the current I_{rf} in Fig. 6 upon the amplitude of the r-f voltage E_{rf} rather than solely upon the direct voltage at the reflector. A small phase shift of this nature is inherent in the bunching process if the r-f voltage is large, and additional phase shifts may result in certain types of reflecting fields. Suppose that the center of a mode is approached from the negative-reflector-voltage side so that oscillations begin with a leading phase shift θ of I_{rf} relative to E_{rf}. Assume that, as the oscillations grow and $|E_{rf}|$ increases, θ changes toward zero. This change of θ favors the supply of power to the oscillations, and a steady-state condition of strong oscillation is obtained. If the reflector-voltage magnitude is now increased, the oscillations continue beyond the point at which oscillations

first began because of the favorable shift in θ provided by the large value of $|E_{rf}|$. Once oscillations begin to die away, they drop to zero amplitude suddenly, because, as $|E_{rf}|$ decreases, θ shifts to values less favorable for power supply. The net effect of this form of hysteresis is that at one edge of the mode reflector voltages fairly near the mode center must be applied to cause oscillations to start but, once started, the oscillations continue for reflector voltages farther removed from the mode center.

The asymmetry indicated in Fig. 5 and the asymmetry and hysteresis produced by a shift in phase associated with $|E_{rf}|$ changes are not particularly serious difficulties from the point of view of electronic tuning. The asymmetries appear in the curves of power output and frequency as functions of reflector voltage, but the power output remains a single-valued, symmetrical function of oscillation frequency. The hysteresis appears only at the edge of a mode, outside the normal tuning range. Much more serious are the *multiple-transit* effects that occur if some of the electrons are permitted to pass through the cavity more than twice.

It would be desirable for all the electrons in the bunched beam returning through the cavity toward the cathode to be collected by the cavity grid near the cathode, the cavity walls, or the accelerating grid. Unfortunately, a few of the electrons continue toward the cathode and, having insufficient energy to reach the cathode, are reversed in direction of motion by the electric field and then are sent through the cavity a third time. Additional bunching effects occur in the cathode reflecting field, and the third-transit electrons provide conditions equivalent to operation in a mode of very high order. The phase of the third-transit bunches varies greatly with tiny changes of reflector voltages, and therefore multiple-transit electrons add "ripples" to the normal curves of power output and frequency. Associated with the ripples are hysteresis effects and frequency discontinuities throughout the mode. Because such effects make automatic frequency control impossible, every effort is made in tube design to minimize the number of electrons that make more than two transits. Sometimes a spike is placed in the center of the reflector electrode to distort the reflection field in such a way as to spread out the returning beam of electrons.

Hysteresis effects and other tuning irregularities may be caused not only by electron motions within the oscillator tube but also by the nature of the load connected to the tube. Discussion thus far has assumed the load to be a resistance R_L shunted directly between the cavity grids in the manner of Fig. 6. Actually, the load is connected through a transmission line and coupling loop to the cavity. Reactance of the coupling loop may cause a dependence of oscillation frequency upon the load or may cause sudden jumps of frequency from one value to another, with a result very similar to that caused by multiple-transit hysteresis. The

effect of coupling reactance is similar to that obtained in triode oscillators, discussed in Art. 10, Chap. X. A long transmission line between the coupling loop and the load may cause similar effects if standing waves are present, as also may reactive elements in the load or discontinuities along the line.

To avoid undesirable effects of the mixer circuit upon the local-oscillator frequency, the device that provides loose coupling between the oscillator and the crystal usually includes an attenuator. For example, in a coaxial-line system, the line from the oscillator may be terminated in a resistance disk matched to the characteristic resistance of the line. The center conductor extends through the disk as a probe to supply local-oscillator power to the crystal, but the majority of the klystron power is dissipated in the resistance disk, and the klystron is unaffected by the small current delivered to the probe. Typical mixer-circuit designs are considered in detail in Art. 16, Chap. XII.

B. MAGNETRONS

The name *magnetron* is applied to tubes of several rather different types, all of which make use, as the name implies, of a magnetic field in the space through which electrons move. The particular type of magnetron employed in radar transmitters is often called a *resonant-cavity magnetron*, and this type alone is discussed here.[1] Radar magnetrons have the basic form of a cylindrical diode, the magnetic field being applied in a direction parallel to the axis of the cathode and anode cylinders. The anode is made from a copper block, and its internal surface is separated into a number of segments by holes in the block that function as cavity resonators. Because radar magnetrons must produce power in short pulses of tremendous amplitude, the tubes are designed for high voltage and high cathode emission. Cooling requirements are reduced because of the long idle periods available between pulses.

5. Magnetron Construction and Type Numbers. The general features of radar magnetrons may be explained by reference to a particular example. A typical magnetron for use in 10-cm systems is the Type 2J32, shown in a cutaway view in Fig. 7*A*. The magnet used with this tube is pictured in Fig. 7*B*. In use, the metal portion of the tube is fitted between the pole pieces of the magnet in such a way that the mag-

[1] For brief descriptions of other types of magnetrons and for references concerning them, see J. B. FISH, H. D. HAGSTRUM, and P. L. HARTMAN, "The Magnetron as a Generator of Centimeter Waves," *B.S.T.J.*, **25** (April, 1946), 167–348 (pp. 177–182 discuss the various magnetron types), or G. B. COLLINS, *Microwave Magnetrons*, 1–8 (McGraw-Hill Book Company, Inc., New York, 1948). The remaining pages of these references contain detailed information on the resonant-cavity magnetron. In *Microwave Magnetrons*, Chaps. 1, 7, 8, 9, and 19 are probably of greatest interest for a first reading.

netic field has the direction indicated by the arrow ℬ in Fig. 7*A*. The cutaway view shows the cylindrical cathode concentric with the central hole in the anode block, and the resonant cavities surrounding the central part of the tube are also visible. A coupling loop in one of the cavities connects to the inner conductor of a coaxial line from which r-f output power is obtained. Radial leads from each end of the cathode assembly permit connection to the cathode and heater. A heavy glass support for the heater-cathode terminals prevents mechanical damage to the cathode assembly. Ordinarily the assembly, consisting of anode, cooling fins, and output-line outer conductor, is grounded, and high-voltage negative pulses are applied to the cathode lead.

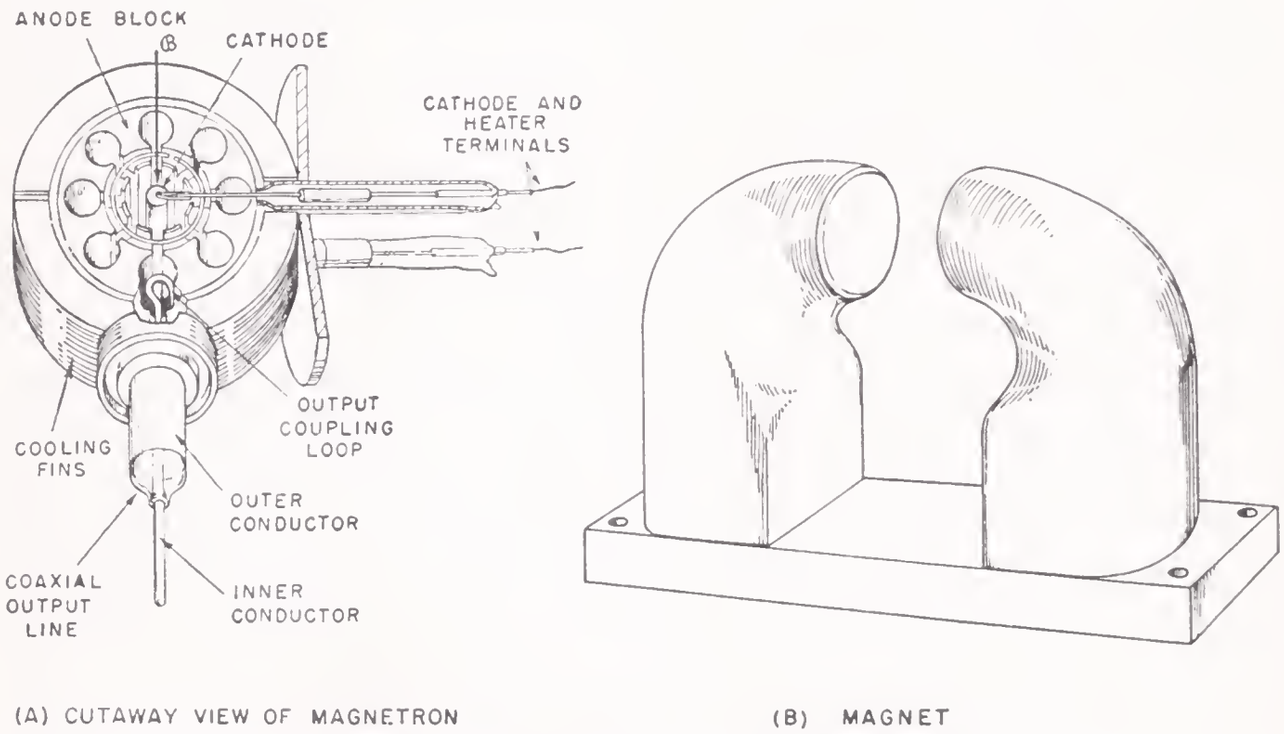


FIG. 7. Type 2J32 magnetron and magnet.

The hole-and-slot form of the resonators in the anode block is shown more clearly by the sectional drawings of Fig. 8. Additional details of the magnetron construction may also be observed in this figure. The anode *straps* comprise metal rings located in slots at each end of the anode block and connected to alternate segments of the block. As explained in detail in the next article, strapping of the anode contributes in an important way to the efficiency of the magnetron. The disks at the ends of the cathode serve to confine the electrons to the interaction space between cathode and inner surface of the anode. The quarter-wave chokes on the heater leads prevent the escape of r-f power to the heater and modulator circuits, because the nearly infinite input impedance of a quarter-wave sleeve appears in series with each heater lead.

The magnetron is comparable to either a triode oscillator or a klystron in that it contains (1) a frequency-determining resonator and (2) an interaction space in which electrons move through the r-f field of the

resonator, contribute energy to the resonator, and sustain the oscillations. In the magnetron, the resonator takes the somewhat complicated form of the anode block in Fig. 8, and the motion of electrons is influenced not only by electric fields caused by the direct anode-to-cathode voltage and the r-f resonator voltage but also by the magnetic field parallel to the axis of the cathode. The nature of the anode-block resonance is dis-

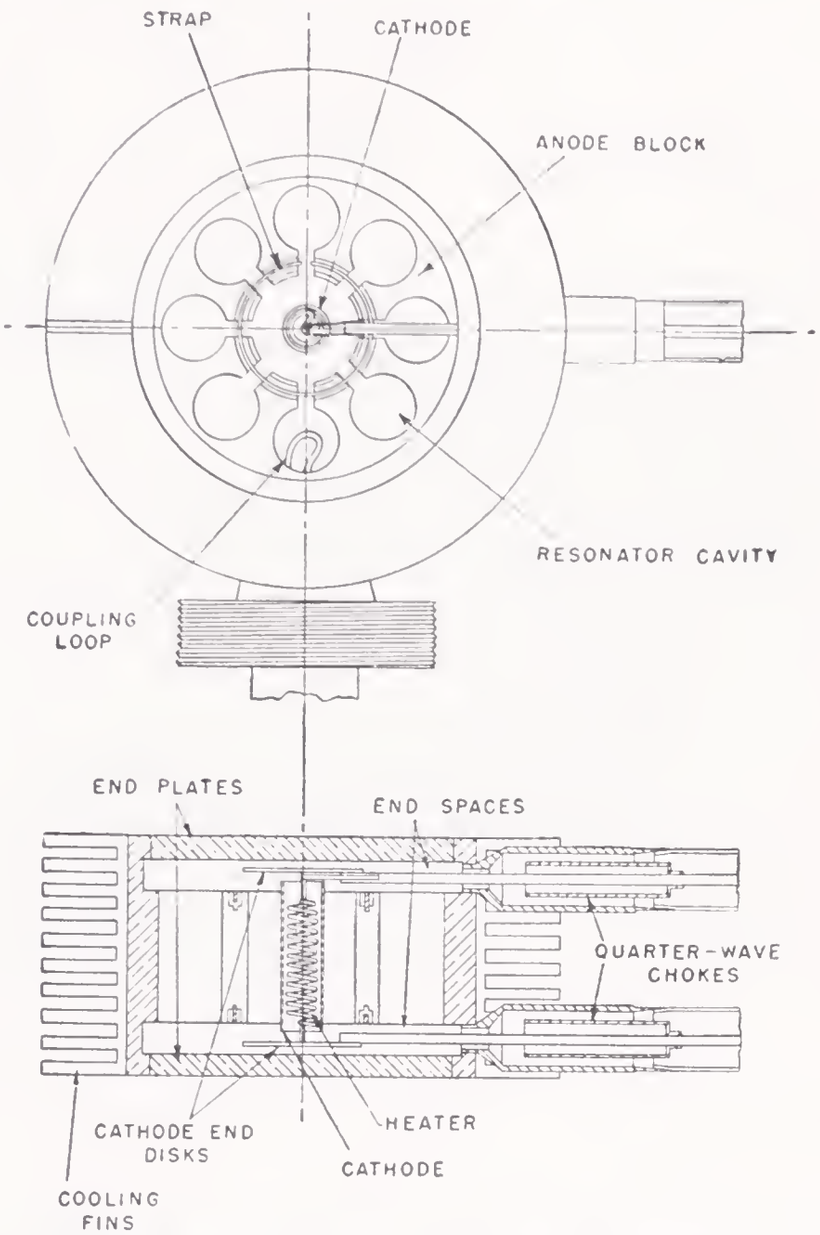


FIG. 8. Sectional views of Type 2J32 magnetron.

cussed in Art. 6, and the way in which the electrons are made to absorb energy from the d-c field and deliver energy to the r-f field is explained in Art. 7. Before these explanations are made, however, it is appropriate to list in Table 2 the type numbers of a selection of radar magnetrons and to indicate some of the differences in construction found among the various tube types.

Observe that the tubes in Table 2 are arranged in groups. The tubes of each group have nearly identical operating characteristics and have oscillation frequencies lying within the small band assigned to the group.

TABLE 2. RADAR MAGNETRONS

Type No.	*Frequency range, Mcps*	*Typical operating conditions*						*Notes*
		Pulse power output, kw	*Pulse duration, μs*	*Repetition frequency, cps*	*Pulse plate voltage, kv*	*Pulse plate current, amp*	*Flux density, gausses**	
700A-700D	680–720	40	2.0	1,000	12	10	650	Unstrapped
4J42	660–730	30	1.5	2,000	12	9	650	Tunable
728A-728J	900–970	260	1.0	1,000	21	20	1,100	
4J51	900–970	285	1.0	1,000	23	20	1,100	Tunable
4J21-4J30	1,220–1,350	440–470	1.0	1,000	22–23	40	1,200	
5J26	1,220–1,350	600	1.0	1,000	27	46	1,400	Tunable
2J22-2J34	2,700–3,333	200	1.0	1,000	18	20	1,800	
706AY-706GY 714AY 718AY-718EY	2,720–3,320	125–146	1.0	1,000	18	16	1,800–2,100	
2J61–2J62	2,914–3,100	70	1.0	1,000	12	12	1,500	Tunable
4J31-4J47	2,700–3,700	400–1,100	1.0	500	25	45–65	2,900	
720A-720E	2,720–2,890	800	1.0	650	24	53	2,600	
4J70-4J77	2,700–3,700	600	0.8	400	26	58	2,500	Tunable
725A, 730A 2J48, 2J50	8,750–9,405	44	1.0	1,000	12	10	5,400	
2J55, 2J56	9,910–9,405	50	1.0	1,000	12	12	3,350	Packaged
2J51	8,500–9,600	40	1.0	1,000	12	12	3,350	Packaged, Tunable
4J50, 4J52, 4J78	9,000–9,405	110, 280	1.0	1,000	15, 22	15, 27	4,950, 6,900	Packaged
3J31	24,000	46	0.5	1,000	14	14	7,600	Rising sun
3J21	24,000	60	0.5	1,000	15	15	8,000	Rising sun, Packaged

* The *gauss* is a unit in the centimeter-gram-second system of units rather than an mks unit. It is used here, in accordance with convention, to specify the flux density provided by magnetron magnets. The gauss is 1/10,000 as large as the weber per square meter. A flux density of 2,000 gausses, for example, is equal to a flux density of 0.2 weber per sq m.

Each group comprises a number of fixed-frequency magnetrons, as well as one or more tunable magnetrons, the tunable magnetrons being adjustable to cover the frequency bands indicated. To a considerable extent, the construction of the Type 2J32 illustrated by Figs. 7 and 8 is typical of all the fixed-frequency tubes. The modifications required in tunable magnetrons are illustrated by the Type 5J26 of Fig. 9 and the Type 2J51 of Fig. 10. The 5J26 is a high-power, 25-cm magnetron having two circular straps at either end of the anode block. Tuning is accomplished by moving a circular member having a U-shaped cross section into or out of slots in the straps at one end of the anode. This adjustment affects mainly the electric field of the resonator and may therefore be thought of

as capacitive tuning.. The 2J51 is a 3-cm magnetron tuned by moving metal pins into or out of the resonator holes in the anode block. This adjustment affects the magnetic field of the resonator and may be considered to be inductive tuning. A third tuning method, employed in the 4J70 to 4J77 group of 10-cm magnetrons comprises an adjustable external cavity coupled to one of the resonator openings in the anode block.

The magnet of the 2J51 tube of Fig. 10 is an integral part of the magnetron rather than a separate unit like the magnet of the 2J32. The magnetron-magnet combination is often called a "packaged" magnetron, and this designation is used in Table 2. The integral construction permits a saving in magnet weight because a shorter air gap is possible; the

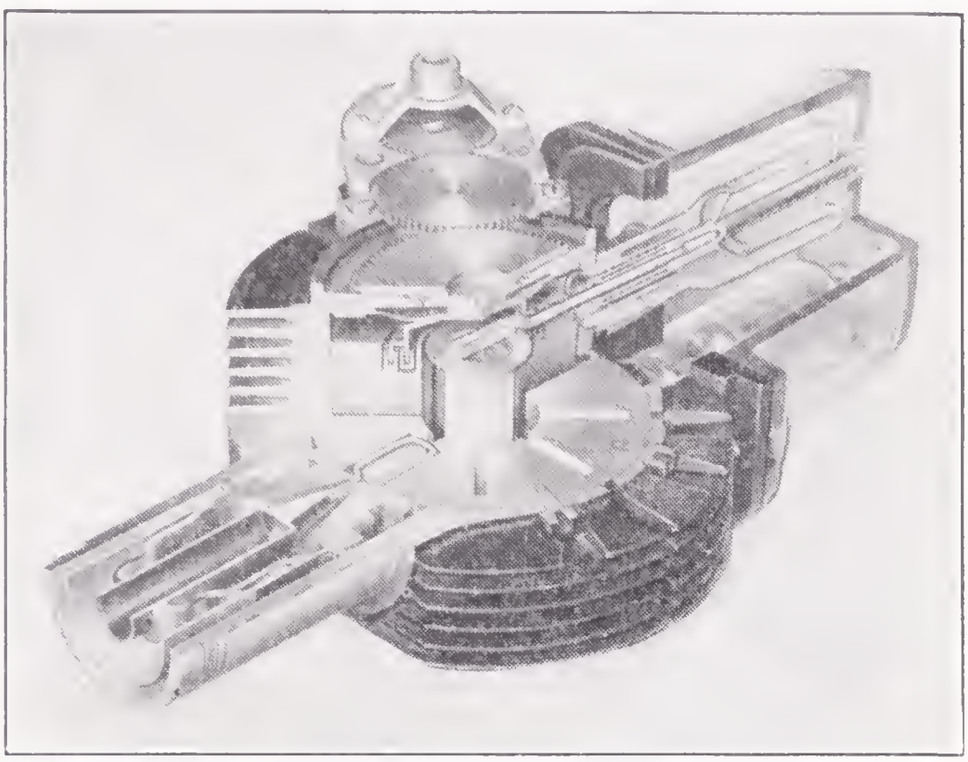


FIG. 9. Cutaway view of Type 5J26 tunable, high-power 25-cm magnetron. (*Courtesy of Bell Telephone Laboratories.*)

advantage of this construction is particularly great with 3- and 1-cm magnetrons because of the high flux densities required by these tubes.

Where a separate magnet is employed, radial leads are required to connect the cathode and heater to the external terminals, as illustrated by the Type 2J32, Fig. 7*A*. Magnetrons are designated as "right-angled" or "straight-through," depending upon whether the radial cathode-heater connections are located 90 or 180 deg from the output connection. The Types 725A and 730A tubes, for example, differ only in that the 725A (see Fig. 21*B*, p. 35, in Chap. I) employs the right-angled construction whereas the 730A is a straight-through type. If an integral magnet is used, the cathode and heater leads are usually arranged coaxially along the axis of the cathode and enter the tube through a hole in the magnet, as illustrated in Fig. 10.

The majority of the magnetrons in Table 2 employ the same general form of anode resonator as does the Type 2J32, though differences in the arrangement of the straps and the shape of the resonator cavities occur from one resonator to another. Exceptions to the use of strapped resonators occur in the 700A to 700D tubes and in the 1-cm tubes, Types 3J31 and 3J21. The 700A to 700D tubes were designed without straps before the technique of strapping was invented. The 1-cm tubes employ the *rising-sun* structure (visible in Fig. 11) in which alternate cavities of the resonator have different dimensions. The manufacture of strapped 1-cm

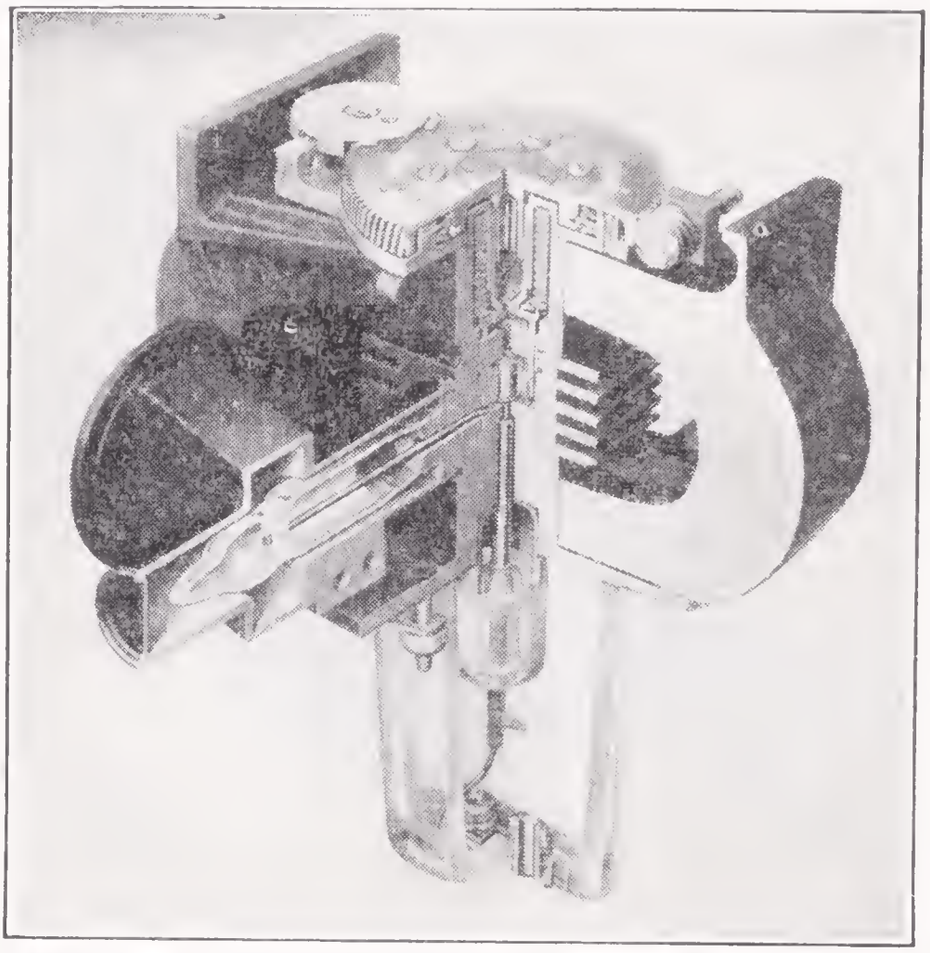


FIG. 10. Cutaway view of Type 2J51 tunable 3-cm magnetron and magnet. (*Courtesy of Bell Telephone Laboratories.*)

magnetrons is extremely difficult because of the small size of the parts. Use of the rising-sun resonator eliminates the need for straps and yields high efficiency in these tubes.

The different output circuits employed in the various magnetrons are of interest. In the Types 2J32 and 5J26 a coupling loop located at the center of one of the resonator cavities is connected to an output coaxial line. A very simple form of coaxial connection is used with the 2J32, and some form of impedance-changing device must be connected between the tube and the output line, if the magnetron is to face its optimum load impedance when the line operates with unity standing-wave ratio. The output system of the 5J26 is more elaborate. A tapered transformer section is included so that the magnetron is *preplumbed;* that is, adjusting

the output line for unity standing-wave ratio automatically provides the optimum load impedance for the magnetron. A choke joint (see Art. 2, Chap. XII) is provided to aid in the transfer of the high power of the magnetron without the losses or sparking that might occur in a simpler junction because of contact resistance.

In the Types 725A and 2J51 3-cm tubes, coupling-loop and coaxial-line output systems are again used. The coupling loops, however, are located at the end of the anode blocks and are sometimes called "halo" loops. These tubes are used in waveguide r-f systems and include probe-type

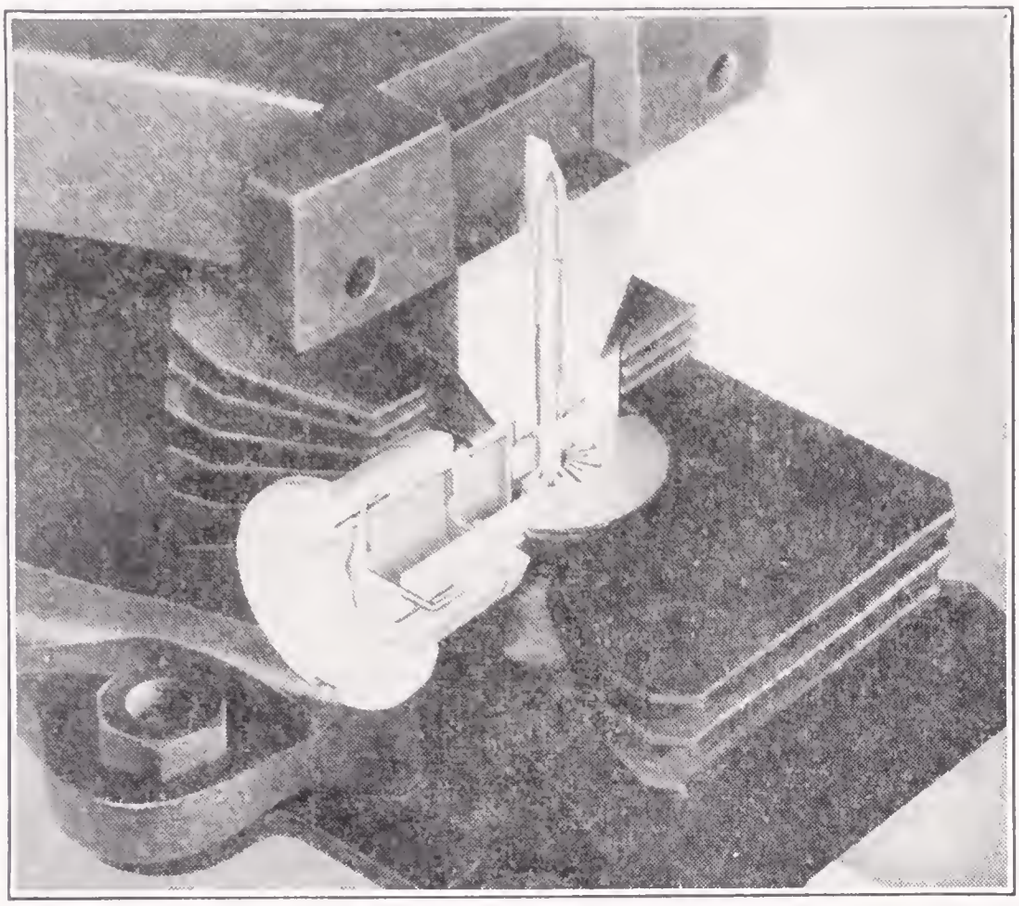


FIG. 11. Cutaway view of Type 3J21 1-cm magnetron and magnet. (*Courtesy of Bell Telephone Laboratories.*)

coaxial-line-to-waveguide transitions. The line-to-guide junctions are designed so that the tubes are properly loaded when the output waveguide operates without standing waves. A simpler method of connecting to a waveguide is illustrated by the 1-cm magnetron of Fig. 11, the waveguide in this tube being coupled by an aperture directly to a resonator cavity. A glass vacuum seal is placed across the guide, and an iris is provided closer to the resonator to permit proper magnetron loading with a flat output waveguide.

6. Magnetron Resonators. The space inside a magnetron, bounded by the anode block, end plates, and cathode assembly (see Fig. 8, p. 765) constitutes a cavity resonator. This resonator is rather different in form from any of those discussed in Chap. IX, but the basic phenomena of resonance—energy interchange between electric and magnetic fields and

energy loss to imperfectly conducting walls—are the same in it as in the resonators of simpler shapes. Of importance to the operation of magnetrons are the mode spectra of magnetron resonators and the field patterns associated with the various modes.

The simplest form of magnetron resonator is the *unstrapped, symmetrical* resonator for which three typical anode-block cross sections are drawn in Fig. 12. The term *symmetrical* signifies that all the cavities in the block are alike. The blocks may contain any even number of cavities. The three parts of Fig. 12 illustrate anode blocks using *hole-and-slot, slot,* and *vane* types of resonator cavities, respectively.

The modes of the magnetron resonator, like those of any other cavity resonator, differ in the complexity of the field-pattern variations along

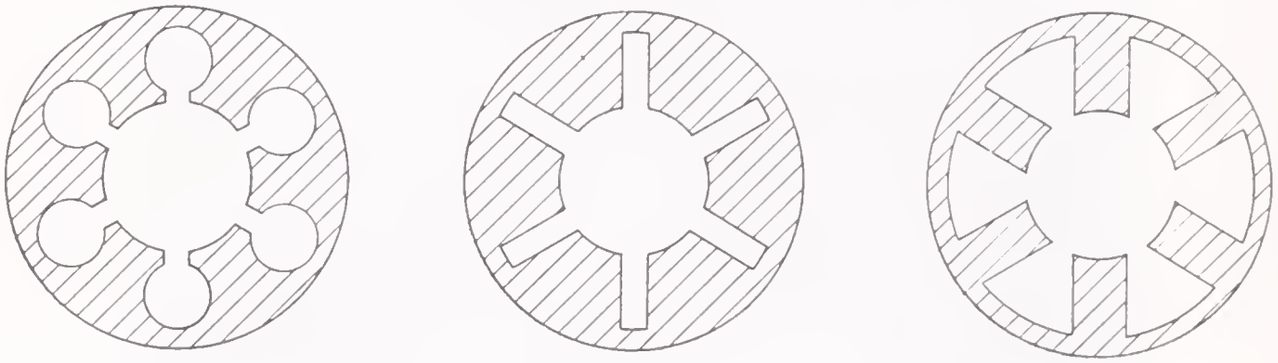


FIG. 12. Cross sections of symmetrical anode blocks.

each of the three axes of a coordinate system. For the magnetron resonator, it is convenient to consider variations (1) along a circle surrounding the cathode, (2) parallel to the axis of the resonator, and (3) along radial lines from the cathode to the outer boundaries of the individual cavities. Ordinarily the only modes of importance are those for which the fields are uniform (except for end effects) along the axial dimension of the anode block and for which the electric field executes a quarter-wave radial variation from zero at the outer boundaries of the individual resonators to a maximum at the inner boundaries. Other modes occur usually in a much higher frequency range. An exception to this statement occurs for magnetrons having closed-end cavities. In these, the end spaces of Fig. 8 are eliminated, and a half-period axial variation of field intensities occurs for the lowest frequency group of modes.

For a four-cavity magnetron having open end spaces, the low-frequency modes have the field configurations sketched in Fig. 13. Actual magnetron blocks contain more than four cavities, but the four-cavity resonator is discussed first because it illustrates the basic principles in a simple manner and can be compared with the four-conductor transmission-line resonators used with ring oscillators (see Art. 5, Chap. X). The modes are identified by the number n, which signifies the number of

cycles of variation of the field quantities that occur around the circumference of a circle surrounding the cathode. Thus, in Fig. 13*A*, the fields are essentially uniform from point to point around the cathode, and the mode designation is $n = 0$. Similarly, in Figs. 13*B* and *C*, one and two cycles of field-intensity variations, respectively, are encountered by a point moving around the cathode, and the respective mode designations are $n = 1$ and $n = 2$. Note that the four metal segments of the anode block may be considered analogous to the four conductors of the transmission-line resonator used with a four-tube ring oscillator. Thus the

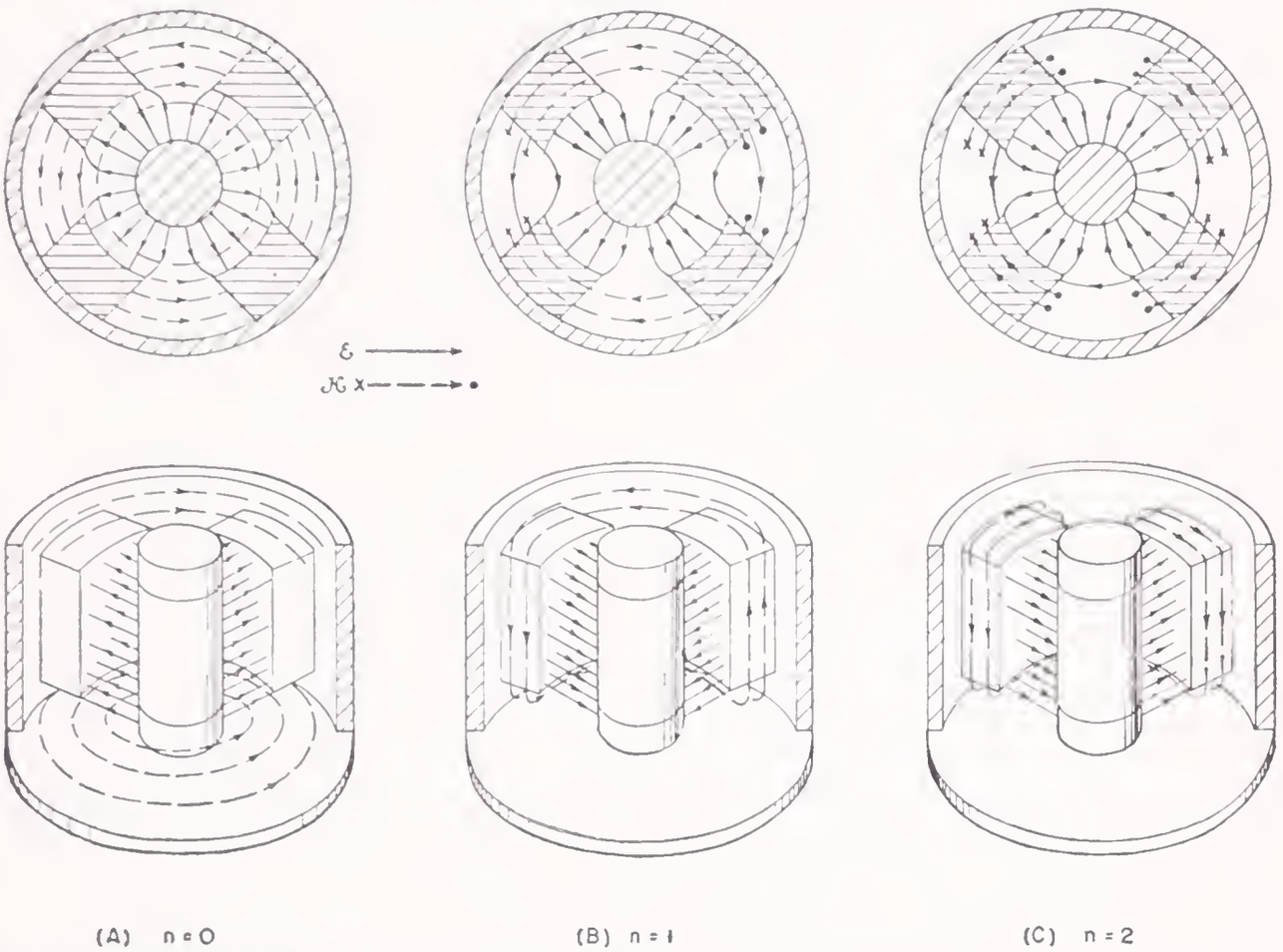


FIG. 13. Sectional views of four-cavity magnetron illustrating field configurations of low-frequency modes.

$n = 0$ mode corresponds to the ring-oscillator mode in which the voltages of all four conductors are the same; that is, the four line conductors are "in parallel." Similarly the $n = 1$ mode corresponds to push-pull-parallel operation in which the voltages of two conductors are in phase and 180 deg out of phase with the voltages of the other two conductors. The $n = 2$ mode corresponds to the ring mode of the four-conductor line in which each conductor voltage is 180 deg out of phase with the voltages of the neighboring conductor. It is this last mode which is the desired, useful mode in magnetrons as well as in ring oscillators. It is often termed the π *mode* because of the π-radian (180-deg) phase change from one anode segment to the next.

For magnetron resonators having more than four cavities, mode designation by names such as "push-pull-parallel" is awkward, but the number n, signifying the number of cycles of variation along a path around the cathode, can always be used to designate the modes. For a magnetron having any even number N of cavities, the modes correspond to $n = 0$, $n = 1$, $n = 2$, . . . , $n = N/2$, and the $n = N/2$ is the desired π mode for magnetron operation. For an eight-cavity magnetron, the $n = 4$ mode is the π mode and has the electric-field pattern in the interaction space (the space between cathode and anode where electrons and fields interact) illustrated in Fig. 14*A*. The pattern of the next-simpler mode ($n = 3$) of the eight-cavity magnetron is shown in Fig. 14*B*.

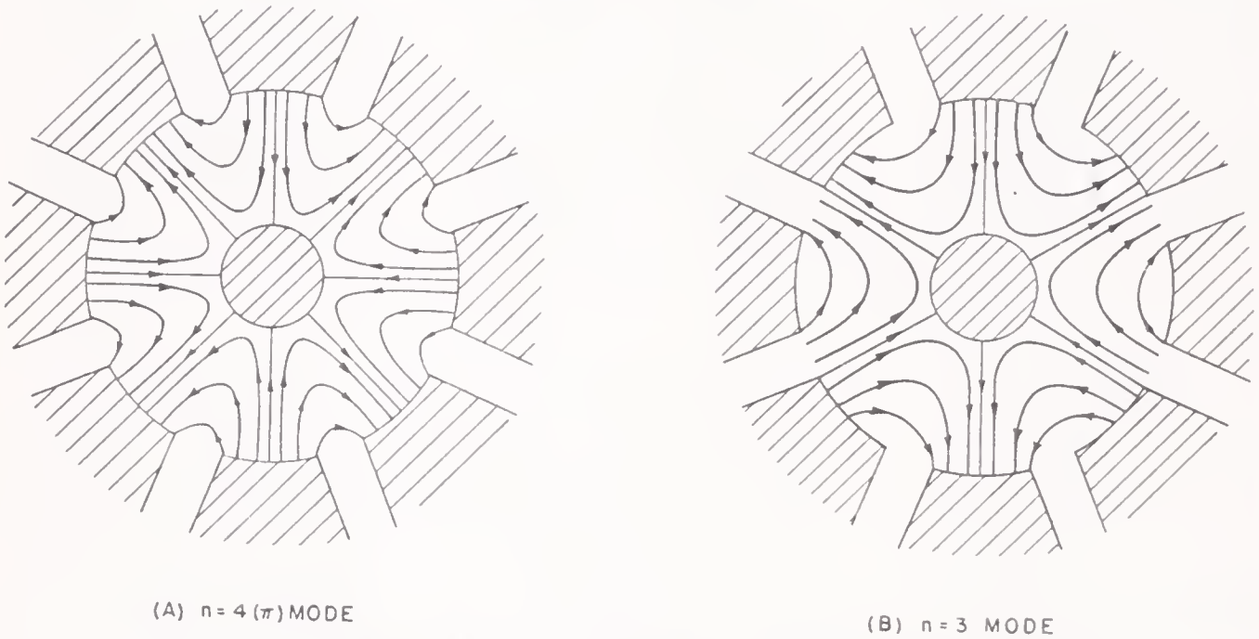


FIG. 14. Interaction-space electric fields of eight-cavity magnetron.

The $n = 0$ mode is not employed in a resonant-cavity magnetron because the fields of this mode do not enter the individual anode cavities, and thus the cavities do not control the frequency of the mode. The other modes, with the exception of the π mode, are undesirable because they are degenerate (see Art. 4, Chap. IX); that is, each is a combination of two modes having identical resonant frequencies. The two component modes are present simultaneously, and the resultant pattern may take a variety of forms depending upon the relative amplitudes of the components. For magnetron resonators, the various patterns of a degenerate mode differ only by a rotation about the tube axis. For example, for the $n = 1$ mode of Fig. 13*A*, the pattern is rotated 90 deg if different pairs of anode segments operate in parallel. A combination of two patterns 90 deg apart yields a pattern shifted by less than 90 deg from the position of the figure. Similar rotations of the $n = 3$ pattern of Fig. 14*B* are possible.

Degenerate modes are undesirable in magnetron operation, for one reason, because the load is ordinarily coupled to only one cavity of the

anode block. A rotation of the field pattern may reduce the fields in the loaded cavity to zero, thus preventing the tube from delivering output power. Another difficulty is that in actual magnetrons unavoidable dissymmetries among the anode segments cause the degenerate modes to separate into two modes of slightly different frequencies. Operation in either of these modes is then possible, but operation is not stable because the slightest change may cause a shift to the other mode. Because the π mode is not degenerate, the position of the mode pattern being fixed by the location of the anode segments, it remains as the desired mode of operation, and in magnetron design provision must be made for preventing the undesired modes from interfering with π-mode operation.

An undesired mode causes interference with π-mode operation if its frequency is very nearly equal to the frequency of the π mode. The movement of electrons in the interaction space is then able not only to sustain the π-mode oscillations but also to produce resonator voltages of lower amplitude in the pattern of the interfering mode. The field in the interaction space is consequently distorted from the π-mode form, and the electron motions are altered in such a way that the efficiency of the oscillator is decreased. This effect occurs frequently in magnetrons with symmetrical unstrapped resonators, because the frequency of the $n = (N/2) - 1$ mode often differs from that of the π mode by less than 2 per cent. Adequate frequency separation of modes cannot be accomplished simply through proper choice of cavity dimensions because other factors impose conflicting requirements on the cavity design. A large anode block is necessary because of the high output power involved, and small cavities are needed in order to produce high oscillation frequencies. The impedance presented to the interaction space depends upon the ratio of the width of segments between adjacent cavities to the width of the cavity slots, and this ratio must be kept reasonably low. Thus a large number of small cavities is required, with the result that undesired modes occur at frequencies very near that of the π mode. However, the frequency separation can be made 10 per cent or more, with considerable improvement in magnetron efficiency, through use of anode straps or a rising-sun type of resonator.

Four methods of strapping are illustrated in Fig. 15. The straps consist of wires of either circular or rectangular cross section connected to alternate segments of the anode block. In single-ring strapping (Fig. 15*C*), all odd-numbered segments are joined on one side of the block and all even-numbered segments are connected at the opposite side. In double-ring strapping (Fig. 15*D*), both inner rings connect to odd-numbered segments and both outer rings join even-numbered segments. The straps may extend beyond the ends of the anode structure as shown in Figs. 15*A* and *B*, or they may be run in shallow slots as in Figs. 15*C*

and *D*. Embedding the straps in the block partially isolates the straps from the interaction space and thus minimizes the distortion in the interaction-field pattern caused by the presence of the straps.

The way in which straps influence magnetron operation may be explained in terms of the additional reactive elements they introduce into the resonator system. Because each strap can carry conduction current between the segments joined, the strap has the effect of connecting the two segments through an inductance. Since a displacement current can also flow between a strap and a segment it passes over, each strap is

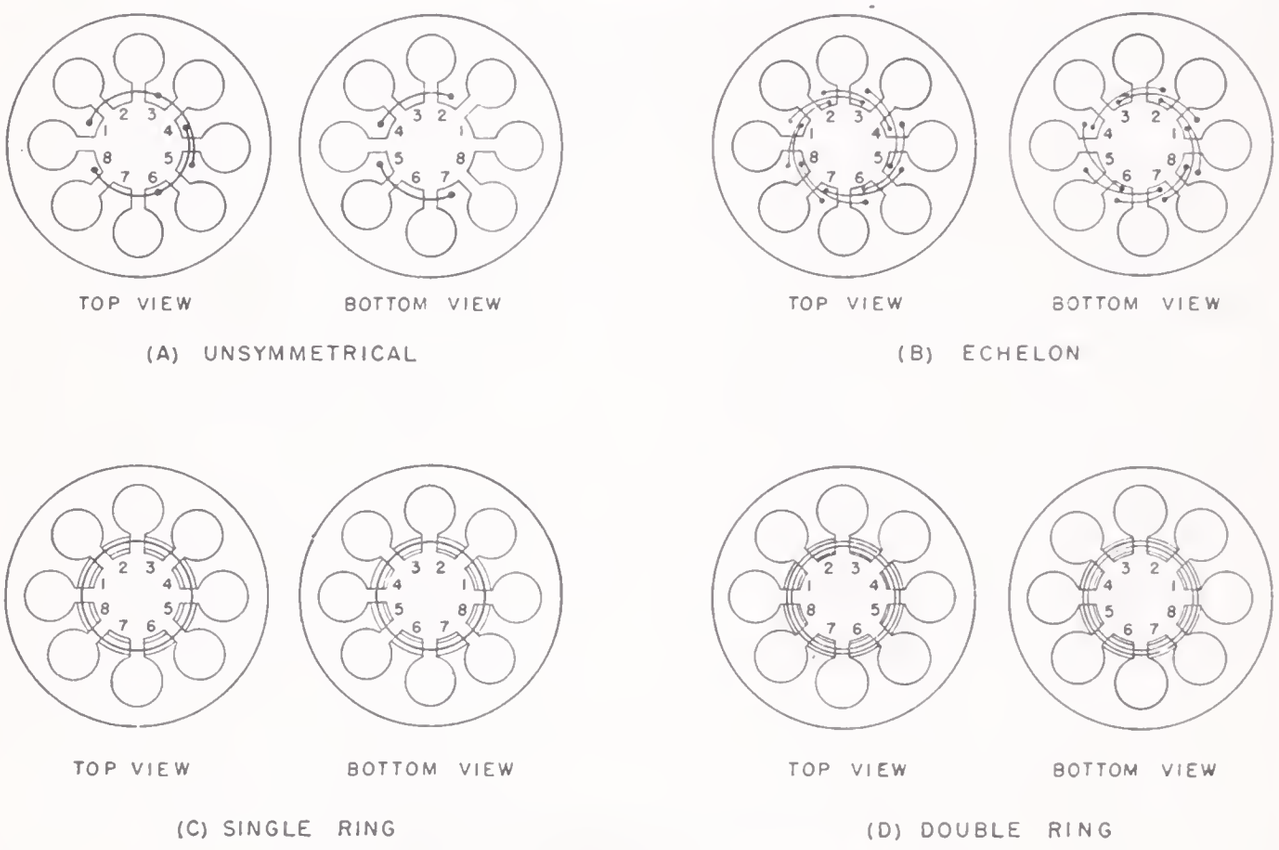


FIG. 15. Methods of strapping.

coupled to a segment through a shunt capacitance. In the π mode, each strap joins segments that are at the same voltage, so that only the shunt capacitance is important. This capacitance is effectively in parallel with the equivalent parallel *R-L-C* circuit of the resonator and hence reduces the frequency of the π mode and lowers the impedance level for this mode. For lower modes the inductance effect of the straps becomes important because for these modes the strap ends are at different voltages. The shunt-capacitance effect is simultaneously reduced because the voltage between a strap and the segment beneath it is less than for the π mode. The combined effect of strap inductance and capacitance is to reduce the resonator inductance, to increase its capacitance slightly, with a net result of an increase in the lower order mode frequencies. As a result of strapping, therefore, the resonant frequencies of all modes are shifted in a manner that increases the separation between the π-mode frequency and the frequencies of all other modes.

The extent to which strapping increases the frequency separation of the modes depends upon the strapping system used. Of the four methods illustrated in Fig. 15, double-ring strapping provides the greatest frequency separation between the π and the next lower mode. Unsymmetrical strapping, on the other hand, yields the least improvement over an unstrapped resonator. Of the remaining two methods, single-ring strapping is the more effective.

Unwanted modes may be further suppressed by introducing an asymmetry into the strapping system. The asymmetry, which may be accomplished by breaking a strap (as in Fig. 15*D*) or by omitting a strap (as in Fig. 15*B*), interrupts current flow in the strap, thereby distorting the patterns of unwanted modes and reducing the efficiency with which energy is transferred from the electron stream to the modes. The amplitude of the oscillations of the unwanted modes is thus reduced. Since the only strap current flowing in the π mode arises from the displacement current between strap and anode segment below, a strap break has very little effect upon the π mode of operation.

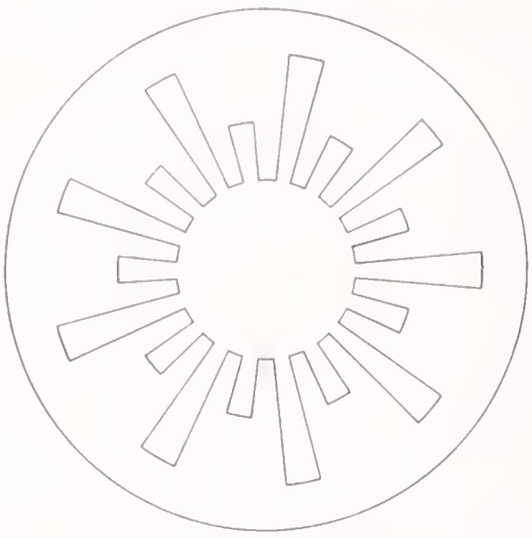


FIG. 16. Rising-sun type of resonator system.

In the absence of strap breaks, degenerate modes may assume any angular position about the resonator axis. Because these modes normally take up a position that yields least coupling to the output circuit, strong oscillations in these modes are possible. The asymmetry resulting from a strap break has the effect of separating degenerate modes into two distinct modes of slightly different frequency. Because the angular positions of the split-mode patterns are determined by the location of the strap breaks, it is possible to locate the output-coupling device so that the split modes are equally loaded. The damping provided by the load then aids in suppressing these modes.

As indicated in Art. 5, strapping is not feasible in 1-cm magnetrons because of the smallness of the anode structure. In the 1-cm region, the desired mode separation is achieved through use of the rising-sun type of resonator, consisting of an equal number of large and small cavities arranged alternately around the anode block (see Fig. 16). The effect of the two sets of resonators is to divide the modes into two groups. As alternate slots are made deeper, the resonant frequencies of modes $n = 1$ through $n = 4$ become less than the frequencies of corresponding modes in the uniform-slot-resonator system. This shift in the mode spectrum

to lower frequencies indicates that oscillations are being controlled mainly by the deep slots. Modes $n = 5$ through $n = 8$ change frequency only slightly, thus indicating that these modes are controlled primarily by the shallow slots. The π mode ($n = 9$ mode) appears in the mode spectrum between modes $n = 4$ and $n = 8$ and is governed by both sets of slots. When the ratio of deep-to-shallow-slot depth is made approximately 2, the π mode is well separated from the others and stable operation of the magnetron is obtained.

7. Electron Motion in Magnetrons. It is evident from the above discussion that multiple-cavity resonators serve to fix the oscillation frequency of magnetrons and hence play the same role in magnetrons as do tank circuits in triode oscillators or simpler cavity resonators in klystrons. This article discusses the manner in which electrons, moving in the interaction space under control of electric and magnetic fields, contribute energy to the r-f fields in the cavities and thus serve to sustain oscillations. This process also has its counterparts in conventional triode oscillators and klystrons. In triode oscillators the resonant circuit obtains energy from electrons moving from cathode to plate of the triode under the influence of the feedback voltage applied to the grid. In reflex klystrons energy is imparted to the cavity by electrons returning through the cavity under the influence of the repeller voltage.

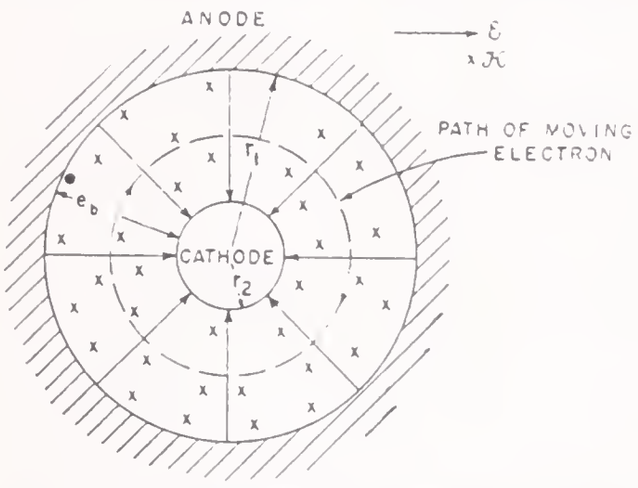


FIG. 17. Cylindrical diode with a uniform axial magnetic field and a direct radial electric field.

In order to gain an understanding of the way electrons move in the interaction space of a magnetron, assume for the moment that r-f fields are absent and that only a uniform axial magnetic field of density $\mathcal{B}$ and a steady electric field of intensity $\mathcal{E}$ are present in the interaction space. The electric field may be considered to result from application of a positive direct voltage e_b between anode and cathode of the magnetron. If fringing fields at the anode segments are negligible, the magnetron becomes, in effect, a cylindrical diode with a magnetic field directed along the cylinder axis and an electric field extending radially inward from anode to cathode. The diode and fields are illustrated in Fig. 17.

Consider, now, the behavior of an electron traversing a circular path at uniform angular velocity in the space between anode and cathode. Such motion is possible provided a mechanism is present to cause the circular motion and provided no resultant force exists to cause the electron to depart from its course, once started upon it. As the electron of mass

m moves along a circular path of radius r with a linear velocity v tangential to the path, it is acted upon by two radially *outward* forces. One is a centrifugal force equal to mv^2/r, and the other, due to the radial electric field, is equal to $Q_e\mathcal{E}$, where Q_e is the charge of the electron. These two forces are opposed by a third force which results because the moving electron constitutes a current in the magnetic field. The third force has a magnitude equal to $\mathcal{B}Q_ev$ and is normal both to the direction of motion of the electron and to the direction of the magnetic field. For clockwise rotation of the electron (counterclockwise current) and $\mathcal{B}$ directed into the page, as in Fig. 17, the force is radially *inward* toward the cathode. Hence, for a given intensity of electric field and angular velocity, it is possible to choose $\mathcal{B}$ so as to balance the radial forces on the electron, thus permitting it to be revolved undisturbed around the cathode. If the electron is revolving midway between anode and cathode under the condition of balanced radial forces, the relationship among forces is

$$\mathcal{B}Q_ev = Q_e\mathcal{E} + \frac{2mv^2}{r_1 + r_2} \qquad (1)$$

where r_1 is the anode radius and r_2 is the cathode radius.

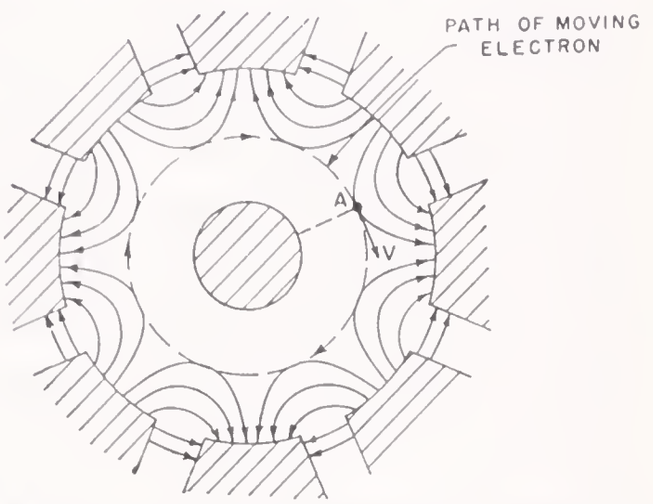


FIG. 18. An electron traveling in interaction space under the conditions assumed for Fig. 17 and in the presence of an r-f electric field.

Suppose that cavities are now provided in the anode structure of the cylindrical diode, and assume that, while an electron is rotating about the cathode in the manner just described, the cavities are excited from an external r-f source that produces the instantaneous r-f electric-field pattern shown in Fig. 18. If the electron arrives at point A beneath an anode gap at the instant the r-f field is at its peak value in the direction of the electron velocity, the field exerts a *retarding* force upon the electron, and consequently the electron gives up some of its energy to the r-f field. Furthermore, if the velocity of the electron is properly related to the frequency of the r-f source, it is possible for the electron to pass beneath each succeeding gap while the field at the gap is in the direction of the electron velocity. As the electron circles the cathode, therefore, it is able to give up energy continually to the r-f field.

It might appear that, because the electron loses speed each time it gives up energy to the field, synchronism between the electron and field would soon be lost. This would be true were it not for the fact that, as the electron slows down, the radially inward force of the magnetic field decreases, thus permitting the steady electric field to urge the electron

toward the anode. The electron is thus able to regain its lost energy from the direct field, and in consequence, it rotates in a larger radius orbit at essentially its initial velocity. Eventually the electron strikes the anode and is removed from the interaction space, but while in the interaction space it may be regarded as the means for transferring energy from the direct field to the r-f field. The external source of cavity excitation originally assumed is not necessary in normal magnetron operation, of course, because any r-f disturbance, such as that due to noise voltages, is sufficient to give rise to minute oscillations in the cavities. These oscillations grow in amplitude at first and are then sustained at constant amplitude by the electron stream.

Actual magnetron performance differs in a number of ways from the idealized operating conditions described above. In order to show how electrons lose energy to the r-f field, for example, it was necessary to assume that the electrons were initially circling the cathode with the correct angular velocity and phase to be retarded by the r-f field. In reality, electrons start almost from rest after being emitted from the cathode, and because the cathode emits electrons uniformly over its surface, there are as many electrons in a position to extract energy from the r-f field as there are to give up energy to the field. On the average, therefore, no net energy transfer would take place were it not that the motion of the electrons is modified as the result of their starting from rest.

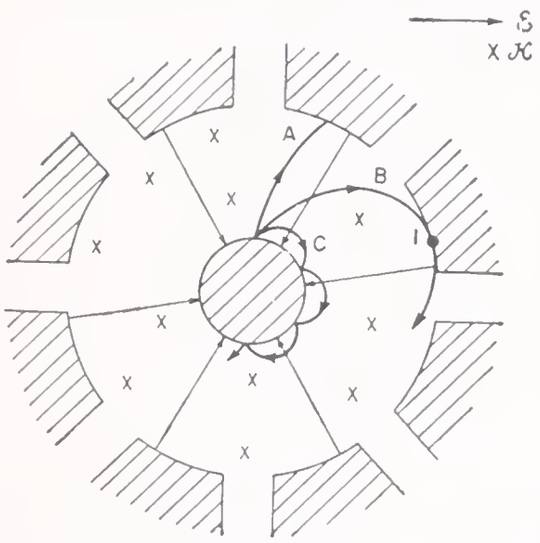


FIG. 19. Possible paths of an electron starting from rest at the cathode of a nonoscillating magnetron.

An electron with zero velocity at the cathode is accelerated radially toward the anode, and as soon as the electron starts to move, the magnetic field exerts a force on the electron normal to its direction of motion. The electron therefore traverses a curved path after leaving the vicinity of the cathode, the amount of the curvature being dependent upon the relative intensities of the electric and magnetic fields. Three possible paths, corresponding to a fixed electric field and three different values of magnetic field, are illustrated in Fig. 19 for a *nonoscillating* magnetron. Curve *A* departs only slightly from a radial line between cathode and anode and is the result of a weak axial magnetic field, curve *B* represents the path for a critical value of magnetic field at which the electron just grazes the anode, and curve *C* corresponds to a strong magnetic field that causes the electron to bend back to the cathode before it reaches the plate. Because the magnetic-field force always acts normal

to the path of motion, it does not affect the energy of the electron. An electron traversing path C loses as much energy to the direct electric field while returning to the cathode as it gained from the field on its flight from the cathode. The electron arrives back at the cathode with zero velocity, therefore, and immediately starts out upon another similar path. As time passes, the electron executes a series of tightly curved paths as it progresses around the cathode.

Observe that in absence of an r-f field either the direct electric field or the magnetic field may be used to control the anode current in the magnetron. For a fixed value of anode voltage e_b, essentially the same anode current flows as the magnetic field is increased from zero to a critical value. At the critical value of magnetic field, the anode current drops sharply to zero and thereafter no current flows as the field is further increased. Alternatively, if the magnetic field is held constant and the anode voltage is increased from zero, no current flows until e_b reaches the critical value that yields curved path B. For larger values of e_b, the current remains constant, since all the electrons are able to reach the anode.

Orbits similar to those of path C in Fig. 19 are of greatest interest in radar magnetrons. This path is very nearly epicycloidal and may be thought of as being generated by a point on the rim of a wheel rolling around the cathode. The resultant motion of the point contains two components—one resulting from the rotation of the center of the wheel about the center of the cathode and the other from the rotation of the point on the rim about the center of the wheel. The first component corresponds to the uniform angular-velocity motion of the electron described in the idealized magnetron of Fig. 17. The second component is superposed on the first and is present because the electrons start initially from rest rather than with a fixed velocity, as assumed in Fig. 17.

When an r-f field is present in the interaction space, the electrons move under the influence of the additional component of field, and consequently they do not traverse the approximately epicycloidal paths described above. The over-all effect of the r-f field is to cause the electrons to *bunch* into groups, the number of bunches depending upon the mode of oscillation and the number of cavities. In an eight-cavity magnetron oscillating in the π mode, the electrons bunch into four uniformly distributed groups, and these bunches circle the cathode at uniform angular velocity. In Fig. 20 concentrations of dots are used to represent the electron bunches. Because of the spokelike appearance of the over-all space-charge pattern, it is sometimes referred to as a "space-charge wheel." The term "space-charge cloud" is also used to describe the pattern.

The reason bunching takes place is that some electrons leave the cathode at times and positions that cause them to be acted upon by an

accelerating r-f field whereas others leave so as to be acted upon by a decelerating r-f field. An accelerating field enables the electrons to extract energy from the field, and in consequence they gain speed. The increase in velocity is accompanied by an increase in the magnetic-field force which causes them to bend back toward the cathode. After one small-radius excursion, these electrons are removed from the interaction space and hence extract no more energy from the oscillating field. Electrons that enter a decelerating field give up energy to the field, and in consequence they lose speed. The decrease in velocity is accompanied by a corresponding decrease in magnetic-field force, and hence these electrons move toward the anode, at the same time regaining their lost energy from the direct field. During their stay in the interaction space, therefore, these electrons serve to extract energy from the direct field and transfer it to the r-f field. As electrons leave the interaction space at the anode after a few oscillations, others enter the formation at the cathode and begin their excursions. The electron bunches are thus preserved.

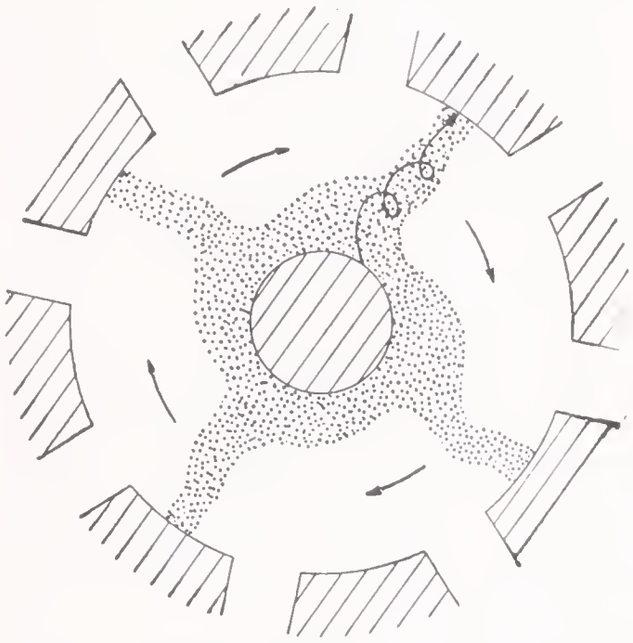


FIG. 20. Representation of the bunches of electrons rotating within the interaction space of an eight-cavity magnetron oscillating in the π mode.

Oscillations can be sustained in the magnetron because the electrons that find themselves in an accelerating r-f field are quickly removed from the interaction space after one short-radius excursion whereas those which encounter a decelerating field are able to remain in the field for a longer period and are able to transfer appreciable energy from the direct to the r-f field before finally reaching the anode. On the average, more energy is transferred to the r-f field from the direct field by the electrons reaching the anode than is extracted from the r-f field by electrons returned to the cathode.

The over-all space-charge pattern shown in Fig. 20 rotates about the cathode in synchronism with the r-f field and in a phase that enables the electron bunches to sweep past a gap at an instant of maximum decelerating r-f field at the gap. Within each bunch, however, electron motions are more complicated, and an exact calculation of the individual paths is rather lengthy and involved.[1] A rough idea of a typical path may be

[1] For a detailed discussion of the interaction of the electrons with the r-f fields of the resonator see G. B. COLLINS, *Microwave Magnetrons*, 207–287 (McGraw-Hill Book Company, Inc., New York, 1948).

obtained if the scalloped orbits of curve C in Fig. 19 are unwrapped from the cathode and then extended from cathode to anode. Actual paths include a series of tight loops of varying radius, as shown by the heavy line within a bunch in Fig. 20.

An effect that preserves the electron bunches and maintains them rotating at proper angular velocity is the *phase-focusing* effect of the r-f field. In Fig. 21 consider an electron passing point A beneath a gap at the instant of a peak decelerating r-f field. An electron at this point is acted upon by the direct radial electric field, the axial magnetic field, and the r-f field, which is tangential to the path of motion. These forces combine to keep the electron rotating at essentially the same angular velocity while it spirals toward the anode, giving up energy to the r-f field in the process. An electron at point B at the same instant is situated so that the r-f field exerts a radial force that opposes the force of the direct field. Its angular velocity is therefore reduced until it falls back in step with electron A. Similarly, an electron at point C encounters a greater radial electric field than the electron at A and consequently rotates slightly faster to maintain the group intact.

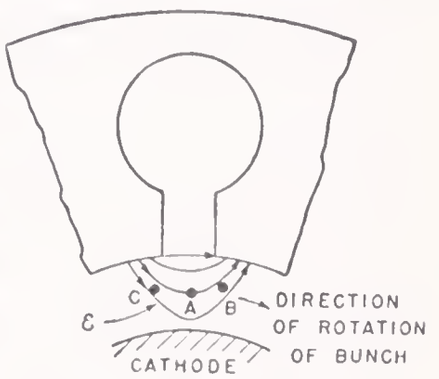


FIG. 21. Diagram for illustrating phase-focusing effect in magnetrons.

Electrons that are excluded from the space-charge wheel return to the cathode and strike it with considerable velocity. Care must be taken, therefore, to prevent damage to the cathode through overheating. Usually, the magnetron heater is connected so that it may be operated at reduced voltage after the magnetron has warmed up.

8. Magnetron Characteristics. Thus far, the magnetron resonator and the behavior of electrons in the interaction space have been discussed separately. It is now desirable to consider the over-all performance of the magnetron and to relate its operation to the concepts presented in the preceding two articles.

An important characteristic of the magnetron is the relation that exists, for various modes of oscillation, between the direct voltage applied to its anode and the density of the magnetic field in the interaction space. This relation is represented graphically by the mode-line diagram of Fig. 22. The diagram is divided into two regions separated by the curve labeled cutoff parabola. The region above the parabola corresponds to the conditions in Fig. 19 that yield the curved path A, that is, to the combinations of e_b and $\mathfrak{B}$ that enable electrons emitted from the cathode to reach the anode in absence of oscillations. The region below the cutoff parabola corresponds to combinations of e_b and $\mathfrak{B}$ that cause electrons in the nonoscillating magnetron to pursue the approximately epicycloidal path C in Fig. 19. In absence of oscillations, no anode current flows in

the region below the cutoff parabola, and in the presence of oscillations, anode current flows only for certain combinations of e_b and $\mathfrak{B}$.

That the boundary between the two regions is a parabola may be shown with the aid of curve B in Fig. 19. This curve represents the critical e_b, $\mathfrak{B}$ combination that causes electrons to graze the anode in the absence of oscillations. Consider point 1 in Fig. 19, the point of tangency of curve B and the anode structure. At this point the radial component of velocity v_r of the electron is zero, and the tangential component v_t is equal to the anode radius r_1 multiplied by the angular velocity $d\theta/dt$, where θ represents angular distance around the cathode. That is, $v_t = r_1 \, d\theta/dt$.

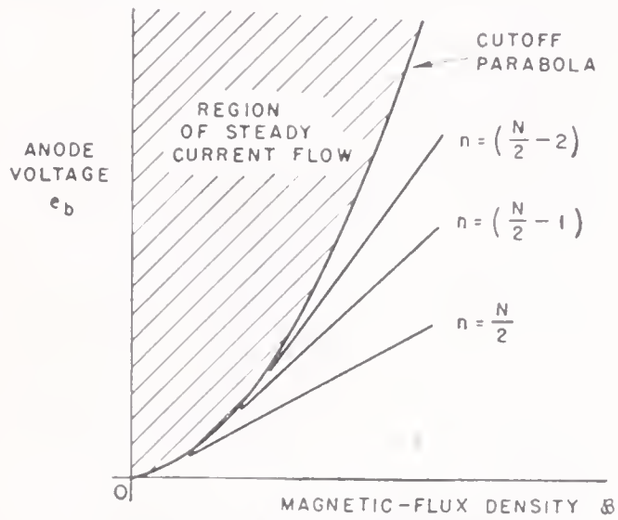


FIG. 22. Mode-line diagram.

If it is assumed that the electron starts from rest, the total kinetic energy gained by the electron in moving from cathode to anode is equal to the potential energy of the electron at the cathode surface, or

$$Q_e e_b = \frac{1}{2} m v_t^2 = \frac{1}{2} m \left(r_1 \frac{d\theta}{dt} \right)^2$$

where, as before, Q_e represents the charge and m the mass of the electron.

In order to introduce the effect of the magnetic field on the electron motion, the relation that torque is equal to the rate of change of angular momentum may be used. At any point along path B and at a radial distance r, the angular momentum of the electron is equal to its moment of inertia mr^2 multiplied by its angular velocity $d\theta/dt$. The rate of change of angular momentum is therefore

$$\frac{d}{dt} \left(mr^2 \frac{d\theta}{dt} \right)$$

At the same point, the torque is equal to the force f_θ acting normally to the radius arm r. This force is the component of the magnetic-field force acting in the θ direction and is due to the radial-velocity component v_r of the electron in the presence of the magnetic field. This force is

$$f_\theta = Q_e \mathfrak{B} v_r = Q_e \mathfrak{B} \frac{dr}{dt}$$

and the corresponding torque is

$$f_\theta r = Q_e \mathfrak{B} r \frac{dr}{dt}$$

Equating the torque to the time rate of change of angular momentum yields the relation

$$Q_e \mathcal{B} r \frac{dr}{dt} = \frac{d}{dt}\left(mr^2 \frac{d\theta}{dt}\right)$$

Upon integration of this expression and substitution of the boundary condition that at the cathode ($r = r_2$) the angular velocity is zero, there results for angular velocity at point 1 in Fig. 19 the relation

$$\frac{d\theta}{dt} = \frac{Q_e \mathcal{B}}{2m}\left(1 - \frac{r_2^2}{r_1^2}\right)$$

If $\mathcal{B}$ has the particular value $\mathcal{B}_{co}$ to yield the cutoff, or grazing, condition, the kinetic-energy expression becomes

$$\frac{1}{2} m \left(r_1 \frac{d\theta}{dt}\right)^2 = \frac{1}{2} m \left[r_1 \frac{Q_e \mathcal{B}_{co}}{2m}\left(1 - \frac{r_2^2}{r_1^2}\right)\right]^2 = Q_e e_b$$

and

$$\mathcal{B}_{co}^2 = \frac{8m}{Q_e r_1^2 \left(1 - \frac{r_2^2}{r_1^2}\right)} e_b \tag{2}$$

The parabolic relation between $\mathcal{B}_{co}$ and e_b is thus proved.

In Fig. 22 the lines indicating modes of oscillations (and hence anode-current flow) beneath the cutoff parabola correspond to combinations of e_b and $\mathcal{B}$ that permit the electrons in the space-charge cloud to circulate about the cathode in synchronism with the r-f fields associated with the cavity oscillations. Several expressions differing by a constant factor may be derived for the mode lines, the exact nature of the expressions depending upon the initial assumptions made in the derivation. One approach is to equate the mean-angular-velocity component of motion of the space-charge cloud to the angular frequency of the r-f oscillations.[1] If the space charge as a whole circles the cathode at uniform speed, it may be assumed that electrons midway between anode and cathode are acted upon by radial forces that are related in accordance with Eq. 1. If the rough approximation is made that direct electric-field intensity $\mathcal{E}$ is equal to $e_b/(r_1 - r_2)$, Eq. 1 becomes

$$\mathcal{B} Q_e v = Q_e \frac{e_b}{r_1 - r_2} + \frac{2mv^2}{r_1 + r_2} \tag{3}$$

In order for these electrons to remain in a decelerating r-f field their angular velocity must equal the angular frequency of the component of

[1] Other approaches and the corresponding expressions are discussed in J. B. FISK, H. D. HAGSTRUM, and P. L. HARTMAN, "The Magnetron as a Generator of Centimeter Waves," *B.S.T.J.*, **25** (April, 1946), 196–197.

r-f field with which they are interacting. Hence, if f_k denotes the component frequency and k the number of cycles of the component field around the cathode, their mean angular velocity ω must equal $2\pi f_k/k$ radians per sec. The corresponding mean linear velocity v is

$$v = \frac{\omega(r_1 + r_2)}{2} = \frac{\pi f_k}{k}(r_1 + r_2) \tag{4}$$

Substituting the expression for v in Eq. 4 into Eq. 3 and solving for e_b yields

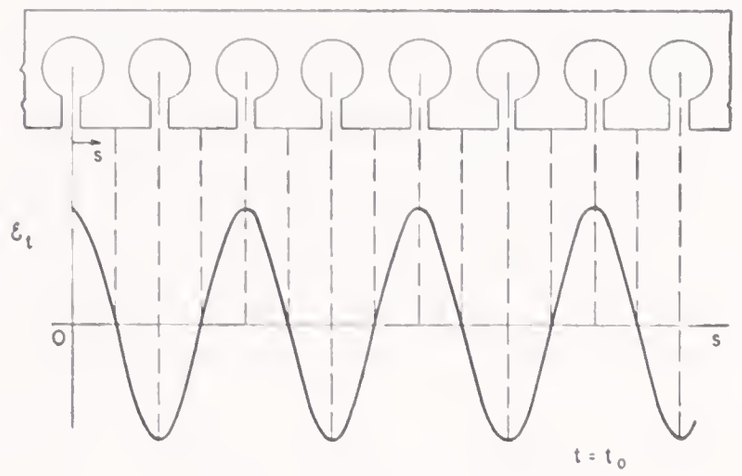
$$e_b = \frac{\pi f_k}{k} r_1^2 \left(1 - \frac{r_2^2}{r_1^2}\right) \mathfrak{B} - \frac{2m}{Q_e}\left(\frac{\pi f_k r_1}{k}\right)^2 \left(1 - \frac{r_2^2}{r_1^2}\right) \tag{5}$$

According to Eq. 5, the mean angular velocity of the space-charge cloud is in synchronism with the r-f field of a mode only for the specific combinations of e_b and $\mathfrak{B}$ that satisfy the equation. It is also evident from Eq. 5 that for any integer k, $\mathfrak{B}$ and e_b are linearly related. Thus if $k = n = N/2$, the locus is the lowermost straight line in Fig. 22.

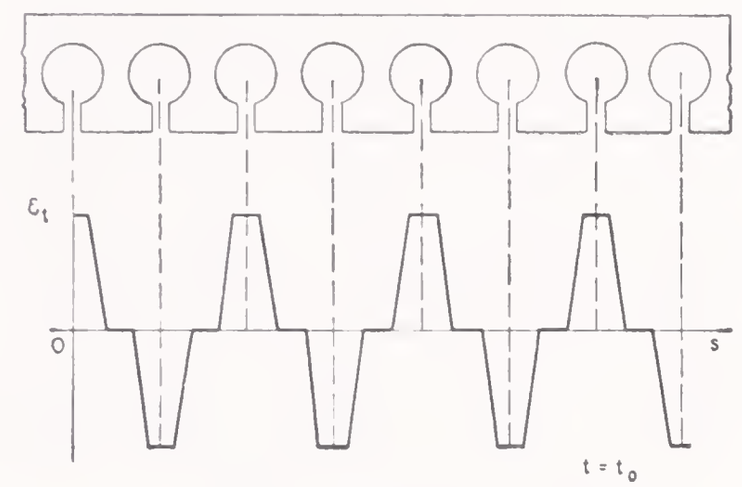
In terms of Fig. 22, the magnetron might be expected to perform in the following manner as its anode voltage is increased from low values and the magnetic field is held constant at a moderately high value. For low values of e_b, anode current is cutoff and no oscillations are obtained. When the $n = N/2$ line is reached, oscillations in the π mode build up and anode current flows. At this point the space-charge cloud circles the cathode relatively slowly, since $N/2$ periods of oscillation are available for the electrons to make a complete revolution. A higher value of e_b causes the electrons to rotate too fast for synchronism with the resonator fields, and oscillations cease. Still higher values of e_b yield synchronism for the next lower mode number $n = (N/2) - 1$. In this mode the space-charge cloud revolves about the cathode faster than before, since it must make a complete revolution in one less period. As e_b is further increased, alternate regions of no oscillation and oscillations in successively lower mode numbers are encountered, until e_b is so large that the region of steady anode current is entered. All possibility of oscillations is then lost. Observe that for very low values of $\mathfrak{B}$, no oscillations are possible either, the cutoff parabola being reached before synchronism with an oscillatory mode is obtained.

Although the magnetron performs in a general way in the manner just described, in actual operation it is found (1) that modes can be excited for combinations of e_b and $\mathfrak{B}$ other than those which determine the mode lines in Fig. 22 and (2) that oscillations can be sustained in a region about a mode line and are not limited to points exactly on a line. Because of these effects overlapping of modes may be observed, more than one mode being possible for a particular combination of e_b and $\mathfrak{B}$. The first effect

may be explained in terms of Fig. 14A, which shows the r-f field pattern of an eight-cavity resonator operating in the π mode. Near the cathode the tangential component of r-f field $\mathcal{E}_t$ varies nearly sinusoidally with angular position about the cathode. If an electron in the interaction space follows a circular path close to the cathode, therefore, it reacts with this sine variation of field. If the electron follows a path near the anode, a tangential electric field in the form of a series of nearly rectangular



(A) $\mathcal{E}_t$ IN REGION OF CATHODE



(B) $\mathcal{E}_t$ IN REGION OF ANODE

FIG. 23. Approximate distribution near anode and cathode of tangential component of r-f field of an eight-cavity magnetron oscillating in the π mode.

pulses would be encountered. The pulse-shaped field near the anode results from the alternations of gaps and segments around the anode structure. Opposite the segments $\mathcal{E}_t$ is essentially zero; opposite the gaps it is either maximum or minimum. Figure 23 shows the eight-cavity anode structure in developed form and the approximate forms of the $\mathcal{E}_t$ waves in the vicinity of the cathode and anode.

The pulselike waveform of Fig. 23*B* can be resolved into Fourier components consisting of a fundamental component and *odd* harmonics. Even harmonics are absent because the symmetrical anode structure yields uniformly spaced positive and negative pulses of equal amplitude.

Oscillations obtained when operation is on the $n = N/2$ line in Fig. 22 correspond to rotation of the space-charge cloud in step with the *fundamental component* of the $\mathcal{E}_t$ wave in Fig. 23*B*. Oscillations in the $n = N/2$ mode may also be sustained if the cloud rotates at one-third the angular velocity required for synchronism with the fundamental component, for the electrons are then able to keep in synchronism with the third-harmonic component of the $\mathcal{E}_t$ wave and deliver energy to the mode through this component. Similarly, if the electrons travel at still slower discrete speeds, they will be able to excite the mode through reaction with the fifth harmonic, seventh harmonic, and so on.

Excitation of the π mode of an eight-cavity resonator is ordinarily accomplished by electrons traveling at an angular velocity corresponding to the frequency of the fundamental component of the r-f field, excitation through interaction with higher order components being rather inefficient because of the relatively low amplitudes of these components. It is quite possible, however, that electrons, while interacting with the fundamental of the π mode, may be rotating at a speed that allows them to interact with a higher order component of another mode. Energy may thus be imparted to the second mode, and the modes may exist simultaneously. Consider the $n = 3$ mode of an eight-cavity resonator (see Fig. 14*B*). Electrons circling near the cathode encounter essentially a sinusoidal distribution of tangential r-f field as before, but the field now comprises three cycles instead of four. The upper sketch in Fig. 24 represents the distribution of the $\mathcal{E}_t$ field near the cathode at a given instant. At the same instant, electrons in the vicinity of the anode encounter a pulse-type distribution of $\mathcal{E}_t$, but as indicated in the lowermost sketch of Fig. 24, the pattern is more complicated than the corresponding π-mode pattern. Because the anode block has the effect of reducing the tangential component of the r-f field to zero opposite the anode segments and of concentrating the field into the regions of the gaps, the resultant $\mathcal{E}_t$ pattern near the anode may be thought of as being formed by multiplying the sine-wave distribution near the cathode by a series of pulses, one for each gap in the anode structure. From this viewpoint the anode block may be considered to have a pulse type of modulation effect upon the sine-wave distribution of field, the shape of the modulation function being given by the center waveform of Fig. 24. Multiplying the modulation function by the sine-wave distribution curve at the cathode yields the distribution curve of $\mathcal{E}_t$ at the anode.

Relative values of the frequency components of the $\mathcal{E}_t$ waveform near the anode may be obtained by determining the spectrum that results when the sine wave near the cathode is modulated by the rectangular modulation function associated with the anode structure. As indicated in Fig. 24, the $\mathcal{E}_t$ field near the cathode recurs at the rate of 3 cycles for

each revolution of a point about the interaction space, and the rectangular pulses of the modulation function recur at the rate of 8 pulses per complete revolution of the point. The rectangular pulses may be resolved into an average component and harmonic components equal to $m \times 8$ cycles per revolution, where m is an integer. Thus the resultant $\mathcal{E}_t$ field near the anode may be thought to consist of a center "frequency" of 3 cycles per revolution and side-band "frequencies" of $3 \pm m \times 8$ cycles

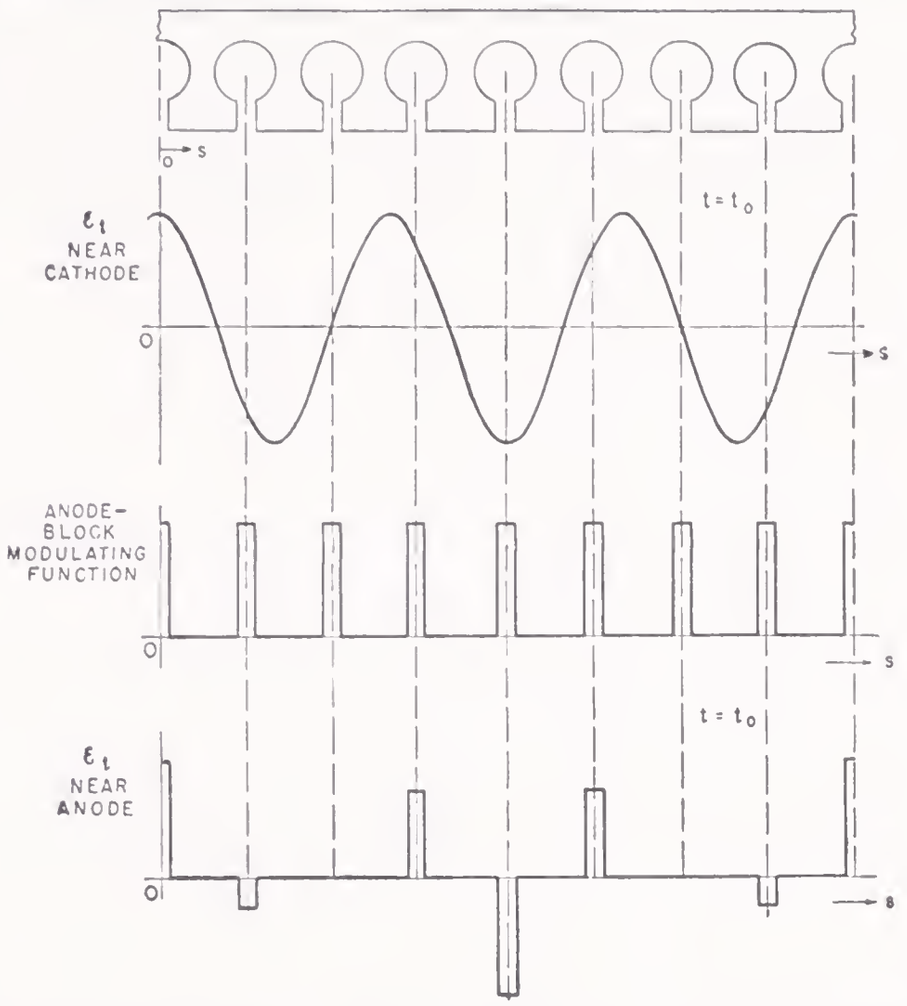


FIG. 24. Approximate distribution near cathode and anode of the tangential component of r-f field of an eight-cavity magnetron oscillating in the $n = 3$ mode.

per revolution. That is, the side-band components are 11, 19, 27, . . . , cycles per revolution and -5, -13, -21, . . . , cycles per revolution.

The above relationships, which apply to the $n = 3$ mode only, may be generalized to include any frequency component of an N-cavity resonator oscillating in any mode number n. If k denotes the number of cycles of the frequency component of the mode,

$$k = |n \pm mN| \tag{6}$$

where $m = 0, 1, 2, 3$, and so on.

It may be concluded from the above discussion that, for any mode number n, the tangential component of the r-f field near the anode contains spatial components that include a fundamental and a number of others. It should be possible, therefore, to excite a mode through reac-

tion of the space-charge cloud with any one of its spatial components. If the mode is excited through its fundamental, the space-charge cloud must rotate at an angular velocity that corresponds to the angular velocity of the fundamental component. For excitation through a higher order spatial component, the space-charge cloud must have a correspondingly lower angular velocity.

Normally only the values of k corresponding to $m = 0$ and $m = 1$ are of importance in mode excitation, the amplitudes of the higher components being too small to yield appreciable excitation. When the space-charge cloud reacts with a spatial component corresponding to $m = 1$ and a value of k equal to $|n - mN|$, the r-f field moves in a direction opposite to that of the fundamental. The space-charge cloud is then said to be exciting a *reverse* mode.

If the frequencies of the $n = 3$ and $n = 4$ modes of an eight-cavity magnetron happen to be in the ratio 5:4, it is possible for the space-charge cloud to excite both modes simultaneously. For this frequency ratio the same angular velocity of space charge is required for exciting the π mode through its fundamental component ($k = 4$) and the $n = 3$ mode through the component corresponding to $k = 5$. Interference of the $n = 3$ mode with the π mode may then be expected.

Sometimes on diagrams like that of Fig. 22, the various mode lines are specified by the symbol $k/n/N$. Thus a line designated as 4/4/8 represents an eight-cavity magnetron oscillating in the $\pi(n = 4)$ mode and excited by electrons reacting with the fundamental component of the r-f field. Although this additional information is often of interest, it should be remembered that the value of n alone specifies the mode. Thus designations such as 3/3/8 and 5/3/8 both specify the $n = 3$ mode but different electron motions for interacting with the r-f field of the mode.

It has been noted that oscillations in a given mode are possible even though the operating point does not lie exactly on a mode line. In the narrow region slightly above the mode line, an effect similar to phase focusing takes place within the interaction space. A slight increase in e_b at a constant value of $\mathcal{B}$ tends to cause the electrons to rotate faster about the cathode, but in so doing they encounter a decelerating *radial* component of r-f field that offsets the excess anode voltage. Oscillations may still be sustained, therefore, despite the slightly high value of e_b. A similar phase-focusing effect also takes place if e_b is made somewhat lower than normal and the operating point falls below the mode line.

Another reason why e_b and $\mathcal{B}$ are not extremely critical is that each electron remains in the interaction space for a time corresponding to only a few r-f cycles. If the electron is not moving in exact synchronism with the r-f field, the efficiency of energy transfer is reduced somewhat but excitation is not entirely lost.

The space charge itself affects the fields within a magnetron and influences operation within a mode. Even in a nonoscillating magnetron operating in the region of no anode-current flow, an appreciable space charge exists in the region of the cathode, and this space charge has the effect of reducing the electric field near the cathode in the same manner as does space charge in an ordinary vacuum tube. Because of space charge, the electric field that is effective in determining the angular velocity of the electrons in normal operation is less than that due to the anode voltage alone. In consequence the magnetron is affected in the following way: Suppose the anode voltage is increased from a low value in the cutoff region until the electrons begin to rotate at a speed that allows oscillations to begin. At this value of anode voltage some electrons reach the anode,

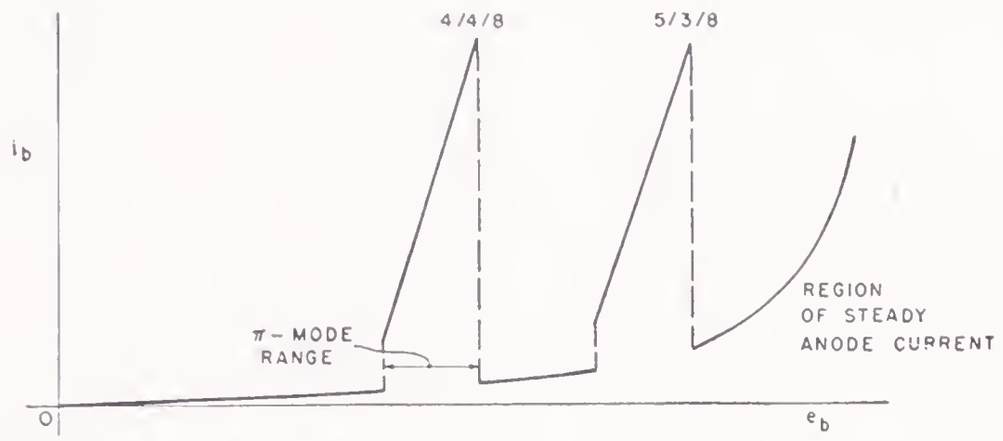


FIG. 25. Typical i_b, e_b characteristic of a magnetron for two modes of oscillation.

the space charge is somewhat reduced, and the direct field is able to draw electrons from the cathode to replace those lost to the anode. A further increase in the anode voltage increases the number of electrons reaching the anode, but at the same time more electrons are drawn from the cathode until the space charge is increased to a point where equilibrium between electrons lost to the anode and electrons drawn from the cathode is again reached. That is, the net effect of an increase in anode voltage in the oscillation region is mainly an increase of anode current. An appreciable portion of the change in the electric field and electron velocity that might normally be expected to accompany a change in anode voltage is offset by the increase in space charge.

The changes in anode current that accompany changes in anode voltage in two modes of oscillation are shown graphically in Fig. 25. The 4/4/8 line is a π-mode oscillation excited through the fundamental, and the 5/3/8 line is an $n = 3$ mode oscillation excited through the $k = 5$ component. A constant value of magnetic-flux density is assumed, and as will be explained in Art. 10, the curves apply for a specific load on the magnetron. Observe that within any mode the i_b, e_b characteristic is essentially a straight line. As e_b is increased, i_b rises slowly as the π mode of oscillations is approached and then jumps sharply as oscillations set in.

Within the mode i_b increases linearly with e_b. Large values of e_b cause oscillations in the $n = 3$ mode, and still larger values push the operating point into the region of steady anode current.

Of great practical importance in radar is the problem of getting the magnetron to operate in the π mode when its anode voltage is supplied in pulses from a line-pulsing or power-amplifier modulator. If the modulator were a zero-impedance source, it would be an easy matter to secure π-mode operation exclusively by selecting a modulator-pulse voltage within the range indicated in Fig. 25. Because actual modulators are not zero-impedance sources, the magnetron behavior is influenced to a great extent by the type of modulator used, and operation in the desired mode is not so easily obtained. A line-pulsing modulator may be regarded as a voltage source with an internal resistance equal to the

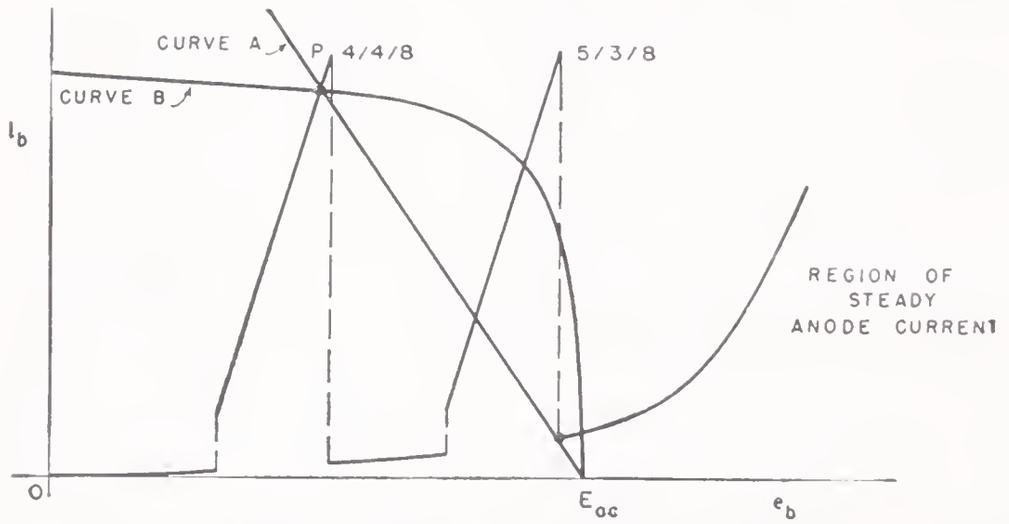


FIG. 26. Modulator characteristics superposed on i_b, e_b characteristic of magnetron.

characteristic resistance of the pulse-forming network (Art. 8, Chap. III), and in normal operation the magnetron is matched to the network through an essentially ideal transformer. Thus the open-circuit modulator voltage E_{oc}, which is equal to twice the magnetron voltage, is applied to the magnetron through a load resistance equal to the internal impedance of the modulator. A modulator load line superposed on the i_b, e_b characteristic of the magnetron thus takes the form of curve A in Fig. 26. For proper operation of the magnetron, the load line should intersect the π-mode line at some point such as P.

In power-amplifier modulators, the amplifier-tube characteristic determines the nature of the load-line curve, and in general the load line is not straight as it is for the line pulsing modulator. Curve B in Fig. 26 is representative of load lines of modulators employing a tetrode in the output stage. The curve may be regarded as a plot of load current and load voltage, for various loads on the amplifier tube. Magnetron operation in the π mode is at the intersection of the load line and the π-mode line as before.

Because the load lines in Fig. 26 intersect not only the π-mode line but also the $n = 3$ line and the cutoff line, operation in both the π and $n = 3$ modes is possible, as is also a condition of no oscillations at all. Magnetrons, like other oscillators, are not observed to operate in their steady state in two modes simultaneously, but they may, and often do, change from one mode to the other, either within a pulse or on successive pulses. A change of modes within a pulse is called *mode shift*, and a change on successive pulses is termed *mode skip*. Where two mode lines lie very close to each other, the prevalence of one mode or the other is determined largely by the relative efficiency with which electrons can react with the r-f fields of the two modes and by the extent to which the modes are loaded. The mode receiving the larger amount of power from the direct field and giving up the smaller amount of power to the load will in general predominate. Where mode lines are well separated, as in Fig. 26, the operating mode is governed mostly by the shape of the anode-voltage pulse applied to the magnetron.

A variety of types of mode skip and mode shift may be encountered in magnetron operation. As an example, suppose a magnetron is being pulsed by a modulator having the load characteristic of curve B in Fig. 26. Operation during the leading edge of the pulse may be visualized by imagining curve B to be displaced initially to the left until its extreme right point coincides with the origin and then moved to the right as the modulator voltage rises. If the voltage increases relatively slowly, the operating point also moves along the π-mode line relatively slowly, and π-mode oscillations start to build up once the steep portion of the π-mode line is reached. Steady-state operation during the pulse is at some point such as P. If the modulator voltage rises rapidly, however, π-mode oscillations may not have time to build up while the anode voltage is in the π-mode range, and since appreciable load current does not flow without oscillations, the modulator voltage would rise above the π-mode range to the region of the $n = 3$ mode. Oscillations may then occur in this mode, or if these also do not have time to build up, the operation point may proceed to the cutoff line.

Erratic operation of this type is most often encountered if high-current, high-power operation within the π mode is attempted, that is, if point P in Fig. 26 is almost at the top of the π-mode line under heavy loading. For relatively low modulator-pulse voltages the π mode may be obtained consistently, but any attempt to increase the voltage applied to the magnetron in order to increase the power output may result in the π mode being skipped on some pulses. A further increase in the pulse voltage may result in consistent operation in the $n = 3$ mode or even in complete cessation of oscillations.

If the leading edge of the modulator pulse is gradual enough to ensure

starting of π-mode oscillations, it is still possible for the magnetron to shift to another mode within a pulse. A mode shift results if the pulse voltage is large enough to cause the operating point to move beyond the upper limit of the π-mode line after oscillations have built up. Because the extent of the π-mode line is diminished by the r-f loading, a magnetron operating normally with one load adjustment may shift modes if loaded heavily. Low cathode emission, which causes low anode current and a shorter π-mode line, may also give rise to mode shift, even with normal load on the magnetron.

The small anode current that flows at very low anode voltages in Fig. 26 is due in part to electrons that leak to the anode by way of the end spaces. The presence of noise in the nonoscillating magnetron also accounts for some of this current. Appreciable noise power is observed in the output circuit just prior to the build-up of oscillations, and it is from the level of this noise that oscillations start in one mode or another.

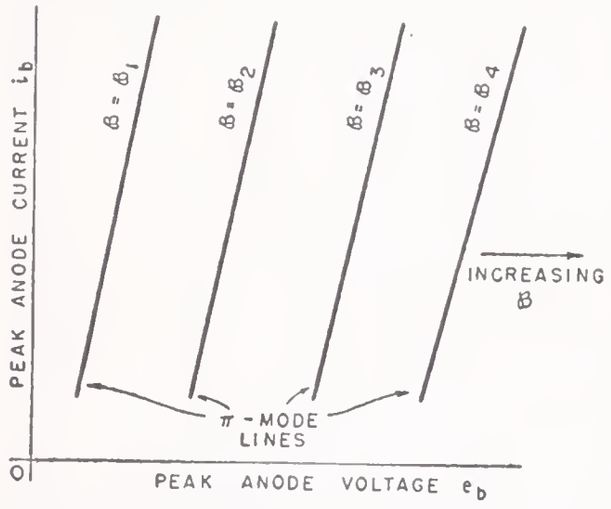


FIG. 27. Basic form of magnetron performance chart for a fixed load.

9. Performance Charts. Curves similar to the i_b, e_b curves of Fig. 26 may be used to determine magnetron performance when the magnetron is pulsed by various types of modulators. In order to include all possible operating conditions, the i_b, e_b curves should be plotted for several different modes and several values of flux densities within each mode. It is evident that such a large family of curves is likely to become very confusing. Because π-mode operation is of primary interest, it is conventional to portray on magnetron performance charts only the i_b, e_b curves that pertain to the π mode. The basic form of such a chart is shown in Fig. 27. Observe that a performance chart does not ensure that the magnetron will always operate in the π mode but merely specifies the magnetron performance when operation is in the π mode. The chart also specifies the operation for only one value of r-f load, and a statement of the r-f loading conditions should accompany the chart.

Although only π-mode lines are included in Fig. 27, regions in which other modes are likely to be excited may be associated with certain areas of the chart. At very low currents the output from the magnetron is essentially noise. Increasing the current slightly causes oscillations to begin, the operating point being at the lower end of the π-mode line. Some tubes have a strong tendency to mode skip or mode shift in this region because of the proximity of other mode lines. For example, it is

possible for an eight-cavity magnetron to have its 5/3/8 line below the 4/4/8 line in the low-current region. The $n = 3$ mode may therefore be excited through reaction of the space-charge cloud with a higher order space component of the mode, and if the mode lines are close together, oscillations may shift back and forth between the two modes.

Stable π-mode operation is most readily secured in the middle range of currents. Even in this region modulator characteristics and r-f loading conditions influence the mode of operation, and these factors must still be considered when the operating point is in the middle range of currents. In the high-current region, excitation of other modes through reaction of

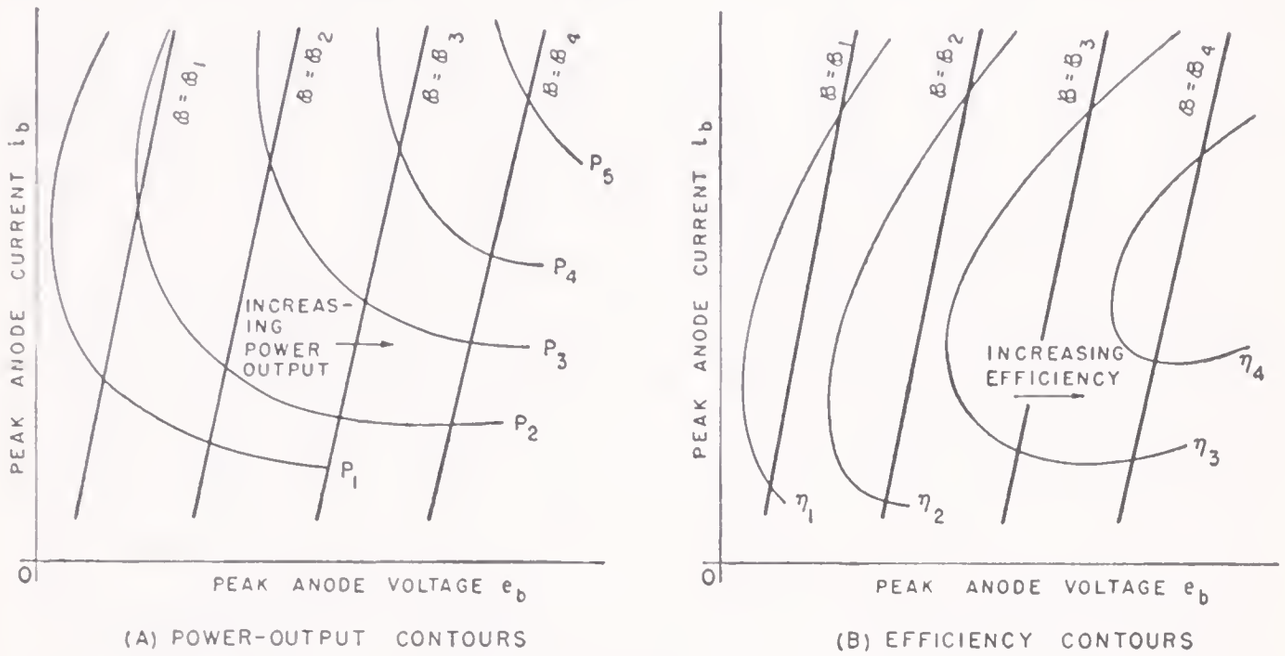


FIG. 28. Power-output and efficiency contours on the performance chart of Fig. 27.

the space-charge cloud with the fundamental of the modes becomes likely, particularly in unstrapped magnetrons. Slight fluctuations in load while the antenna scans are usually sufficient to cause a shift to another mode. In strapped magnetrons, on the other hand, the mode separation is often great enough to permit operation near the extreme end of the π-mode characteristic.

The usefulness of a performance chart may be increased by including additional magnetron properties on it. For example, a set of r-f power-output curves may be drawn, as shown in Fig. 28*A*. These curves are obtained by measuring the r-f power output at constant load for many different operating points and connecting points of equal power. If the magnetron efficiency remained constant, the power output would be proportional to power input, and the power-output contours would be hyperbolas that approach the i_b and e_b axes asymptotically. Actually, the curves deviate from hyperbolas because the efficiency is a function of the operating point. Efficiency contours are shown superposed on the i_b, e_b curves in Fig. 28*B*.

Two effects may be noted from the efficiency curves. One pertains to the performance along a constant flux-density line, and the other to performance at constant anode current. Along a constant flux-density line the efficiency goes from relatively low values at low currents to a maximum value in the middle-current region to lower efficiencies at high currents. The change of efficiency with operating current is related to the extent to which the space-charge cloud departs from synchronism with the tangential component of the r-f field. In the high-current region, the anode voltage exceeds the value that yields optimum synchronism, and electrons, before striking the anode, tend to become slightly debunched and move ahead of the positions for greatest interchange of energy from the direct field to the r-f field. A similar situation exists at low currents, except that the electrons tend to lag. In addition, the low efficiency at low currents is partially the result of energy lost to leakage currents, that is, to electrons that reach the anode by paths other than those which permit energy transfer to the oscillations.

Along a line of constant anode current the efficiency increases steadily from the region of low e_b and $\mathcal{B}$, to the region of high e_b and $\mathcal{B}$. As e_b and $\mathcal{B}$ are increased in a manner that maintains constant anode current, the electrons in the space-charge cloud maintain synchronism with the r-f field while acquiring more energy from the stronger direct field. Despite their acquiring more energy from the direct field, however, the electrons reach the anode with essentially the same velocity, independently of e_b. Thus most of the increased energy is transferred to the r-f field and the efficiency is improved. The failure of the larger anode voltages to increase the final speed of the electrons is associated with the fact that the electrons do not spiral smoothly from cathode to anode but rather execute a number of cycloidallike loops superposed upon the spiral (see Fig. 20). If e_b and $\mathcal{B}$ are increased simultaneously, the larger voltage accelerates the electrons more quickly to the velocity at which the increased magnetic field bends them around the looped path. The effect is mainly to make the size of the closed loops smaller and to maintain the velocity gained in each loop nearly constant. The angular velocity at which the electrons circle the cathode is thus preserved and also the velocity at which the electrons strike the anode.

Usually contours of constant frequency are also added to magnetron performance charts. As for the other contours, the conditions of r-f loading must be specified, and in addition the temperature is important, since cavity dimensions, and hence cavity resonant frequency, are dependent upon temperature. Instead of plotting absolute frequency on the chart, a reference frequency is often chosen, and *frequency changes* from the reference are plotted. A complete performance chart that includes

frequency, power-output, and efficiency contours is given in Fig. 29 for a Type 2J32 magnetron.

10. Rieke Diagrams. The performance charts discussed in the preceding article specify the operation of magnetrons in terms of magnetic-flux density, input voltage, and input current for a specific r-f load. Frequently it is also desirable to know how the magnetron performs when its input parameters are held constant and the load is varied. Diagrams that provide this information are called *Rieke diagrams.*

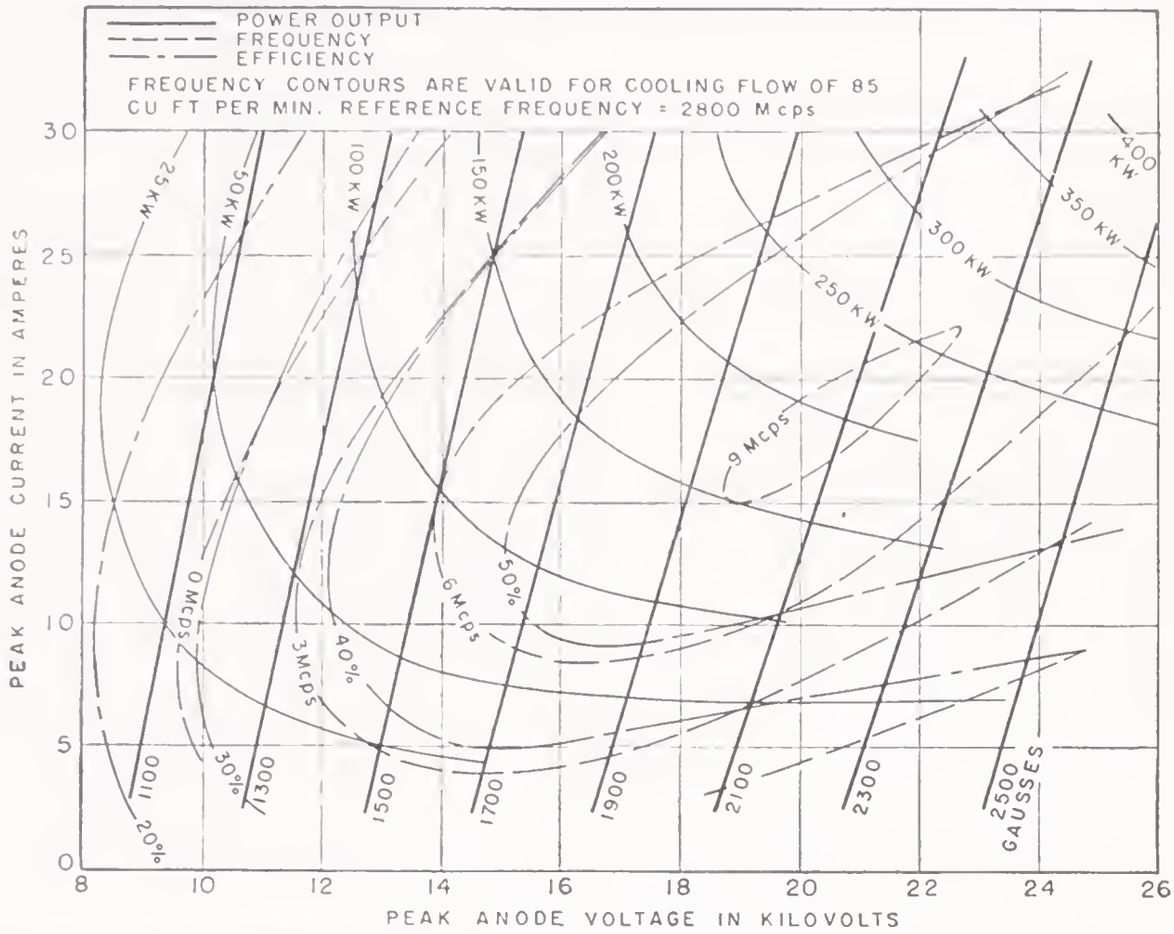


FIG. 29. Average characteristics of a 2J32 magnetron. The magnetron load is a coaxial line with a matched termination.

As a preliminary to the discussion of Rieke diagrams, it is helpful to consider the magnetron and r-f load in terms of an equivalent lumped-parameter circuit. Magnetron oscillators, like klystron and vacuum-tube oscillators, may be represented, for a single mode, as a current source feeding a parallel resonant circuit shunted by an r-f load. In fact, an eight-cavity magnetron operating in the π mode may be thought of as a high-frequency analogue of an eight-tube ring oscillator. The similarities between the two may be explained from Fig. 30. In the magnetron the four groups of electrons in the rotating space-charge cloud deliver electrons to the eight anode segments in much the same way that the eight triodes deliver electrons to their eight plates in the ring oscillator. Thus when the space-charge cloud is in the position shown in Fig. 30*A*,

anode segments 1, 3, 5, and 7 receive electrons and segments 2, 4, 6, and 8 receive none. Similarly, when tubes 1, 3, 5, and 7 in Fig. 30*B* are conducting, tubes 2, 4, 6, and 8 are cutoff. One-half cycle later, the even-numbered anode segments of the magnetron receive electrons and the even-numbered tubes of the ring oscillator conduct. In the ring oscillator an additional set of cathode or grid lines is required, of course, in order to obtain the necessary feedback voltages for sustaining oscillations. The corresponding feedback mechanism in the magnetron is the magnetic field, which causes the electrons to bunch and rotate in synchronism with

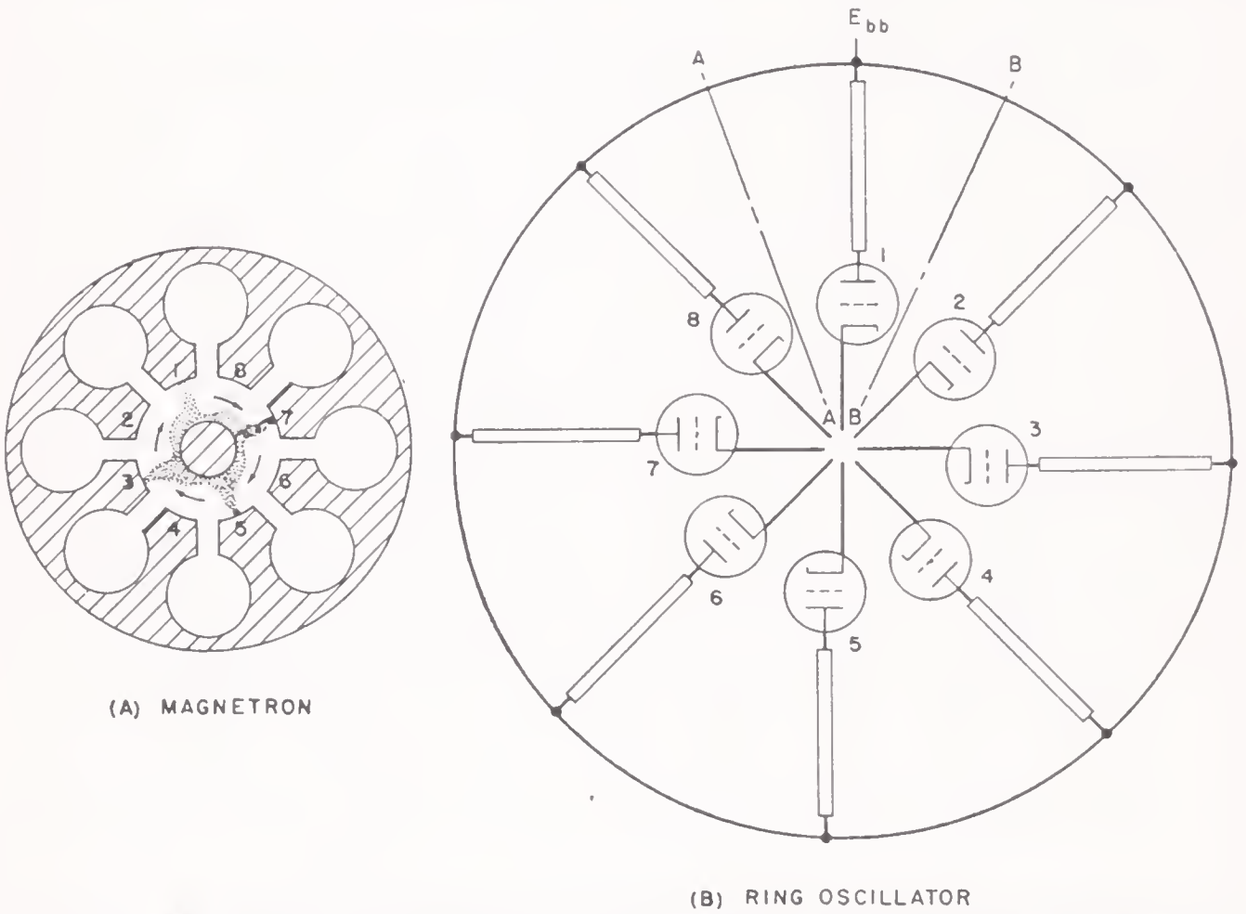


FIG. 30. An eight-cavity magnetron compared with an eight-tube ring oscillator.

the r-f field. A tendency for the groups to pull out of synchronism is equivalent to an advance or delay in the phase of ring-oscillator feedback voltage.

The oscillator of Fig. 30*B* may be reduced to an equivalent single-tube circuit by applying the methods used to derive the ring-oscillator equivalent circuits in Arts. 5 and 7 of Chap. X. With respect to the symmetry planes *A* and *B* in Fig. 30*B*, the single-tube circuit takes the form of Fig. 31. A similar magnetron equivalent circuit may be constructed by imagining perfectly conducting planes of zero thickness to be present along radial lines through the centers of the cavities on either side of segment 1. Locating conducting planes in these positions is possible because, as indicated in Fig. 32*A*, the $\mathcal{E}$ component of the oscillating field is normal and the $\mathcal{H}$ component is tangential to the planes. Observe in

Fig. 32*A* that, whereas the $\mathcal{E}$ field is concentrated near the slots and diminishes along radial lines extending into the cavities, the $\mathcal{H}$ field is weakest at the slots and strongest inside the cavities. Thus the combination of anode segments and two half-cavities bears a definite resemblance to a quarter-wave, shorted-end line resonating in its lowest mode and in an equivalent circuit may be represented as a parallel combination of L and C (see Fig. 32*B*). Energy for sustaining oscillations is derived from the current in the space-charge cloud sweeping past the anode segments and is accounted for in Fig. 32*B* by the complex number I_o standing for the fundamental component of the r-f current supplied to the oscillations. The share of the output load contributed by segment 1 and adjacent half cavities is represented in the equivalent circuit by admittance Y_L coupled to the L-C resonant circuit through an ideal transformer and coupling inductance L_c, the transformer and coupling inductance being introduced because connection is made to the load through a coupling device.

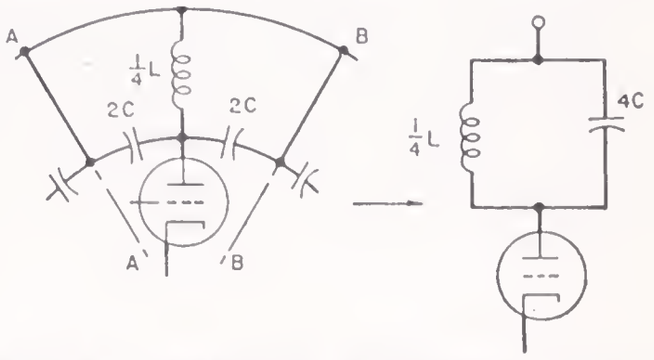
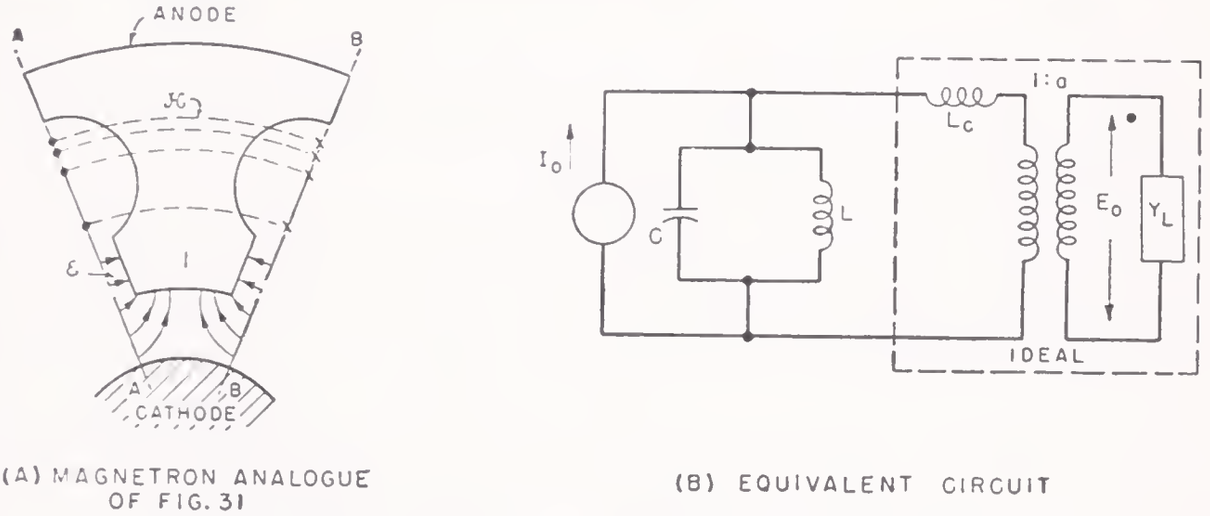


FIG. 31. Single-tube equivalent of ring-oscillator plate circuit of Fig. 30*B*.

Figure 32*B* may be used to explain the effect of load changes on the magnetron operation. Consider first that the coupling reactance is



(A) MAGNETRON ANALOGUE OF FIG. 31

(B) EQUIVALENT CIRCUIT

FIG. 32. R-f equivalent circuit of magnetron.

negligible and the load is a pure conductance. The primary effect of load changes is then on the power output. The frequency remains essentially constant, although slight changes are to be expected because of the changes in space charge that accompany load changes. A small value of load conductance causes the oscillations to build up to the largest value the space-charge cloud can maintain and consequently the load voltage E_o is large. Because of the small conductance, however, the power out-

put is low. Increasing the conductance increases the load current without affecting the amplitude of the oscillations appreciably, and hence the power output increases. Maximum power output occurs when the reduction of E_o that accompanies the increased current and loading is exactly offset by the current increase through the load. Beyond the maximum the power decreases, and eventually, under heavy loading, oscillations cease altogether.

If the load contains a susceptive component as well as conductance, the resonator frequency becomes dependent upon the value of the load susceptance, since the resonator is in shunt with the load. Thus adding capacitive susceptance lowers the frequency, and adding inductive sus-

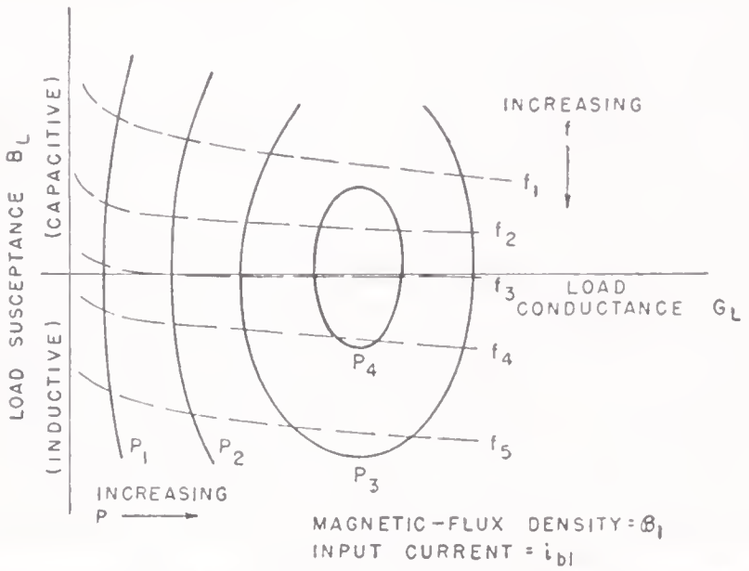


FIG. 33. Magnetron frequency and output-power contours in the load-admittance plane.

ceptance raises it. The output power also decreases slowly with changes in frequency because of the tendency toward loss of synchronism between the space-charge cloud and the r-f field.

These effects are portrayed graphically in Fig. 33, which shows contours of constant frequency f and constant output power P in the load-admittance plane. Observe that, if the load is a pure conductance, the output power goes through a maximum as the conductance is increased and the frequency remains essentially constant. On the other hand, if the conductance is fixed and the susceptance is varied, the frequency is highest for large values of inductive susceptance and lowest for large values of capacitive susceptance. Only moderately small power changes occur over the complete range of susceptance values, but maximum power coincides with zero load susceptance. Diagrams like that of Fig. 33 apply for a specific set of input conditions, that is, for a definite value of magnetic-flux density and input current or voltage. The values of these quantities should be specified on the diagram, just as the load conditions should be specified on the performance charts discussed in the preceding article.

Figure 33 is sometimes referred to as a *Rieke diagram*, but more often this name is applied to a similar plot on a Smith chart, that is, to a diagram that shows frequency and output-power contours in the reflection-coefficient plane. The reflection coefficient is determined by the relation of the load impedance to the characteristic resistance of the transmission line or waveguide attached to the magnetron output-coupling device. Use of the Smith chart has the advantage that operating conditions of the magnetron can be readily determined from standing-wave measurements on the output line. When a transmission line is attached to the output-coupling device, it is important that the reference plane for the load admittance be specified.

The result of transferring the data in Fig. 33 to the reflection-coefficient plane is shown in Fig. 34. Magnetron performance for various types of loads may be obtained from this figure. Assume that the load admittance is the input admittance of an r-f line. If the line is terminated at the far end in a matched load, operation is at the center of the Rieke diagram, and the frequency and power output are those at the chart center. Actually in radar systems an exact match cannot be expected over the entire line because of small reflections introduced by rotary joints, bends, stubs, and small changes of antenna impedance during scanning. The operating point does not remain stationary, therefore, but rather it shifts around within a small region about $g_L = 1$. Consequently, the power output and frequency may be expected to vary continually in small amounts about the value at the chart center.

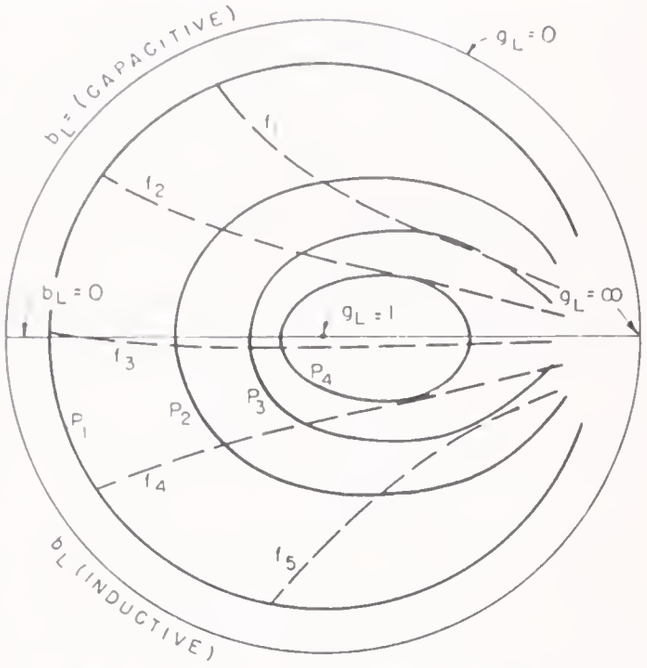


FIG. 34. Basic form of Rieke diagram. The diagram applies at a specific point of the output line connected to the magnetron.

The effect of load changes on magnetron frequency is called *frequency pulling*, and the *pulling figure* denotes the total frequency change that results when the load on the magnetron is varied in a manner that yields all possible positions of the standing-wave pattern on the output line. the standing-wave ratio being held constant at 1.5. Thus the pulling figure is in effect a figure of merit of the tube in that the lower the pulling figure, the smaller is the frequency change when the r-f load is varied in a particular way.

Ordinarily an impedance-changing device is included between the output coupling loop and the r-f line connected to the magnetron. The

effect of the device is to move the operating point away from the center of the Rieke diagram to a point determined by the input impedance of the line section attached to the coupling loop. Suppose, for example, that the input impedance is a pure conductance. The operating point is then shifted along the $b_L = 0$ line either to the right or left of the chart center, depending upon whether g_L is greater or less than unity. From the standpoint of output power alone, operation near the right edge is desirable because of the high-power output obtained in this region. Unfortunately, however, serious frequency instability is likely to result because the frequency contours tend to converge at the right-hand edge. For good frequency stability the operating region should include as few frequency contours as possible and hence should be located in the left-hand portion of the Rieke diagram where the per-unit conductance is less than unity. Because the power output is low in this region, the operating region chosen in practice represents a compromise between good frequency stability and high-power output.

Thus far it has been assumed that the magnetron cavities have been coupled through an ideal transformer to a matched or nearly matched line and that the impedance-changing device is located close to the magnetron. Actually, some coupling reactance may be associated with the output loop, and also it is possible in some magnetrons to locate the impedance-changing device an appreciable distance from the loop. Both of these possibilities can give rise, if adjustment for high power is attempted, to more severe frequency instability than the frequency pulling described above. The instability takes the form of sudden jumps in frequency occasioned by slight changes in r-f load.

Coupling reactance has the same effect upon the magnetron frequency as it does on the oscillation frequency of a triode oscillator. If the coupling reactance of the triode oscillator is series tuned and an attempt is made to load the oscillator heavily, it is found that instead of more power being delivered, the frequency jumps to a new value, and slight changes in load cause the frequency to shift back and forth between two values. (This effect is discussed in detail in Art. 10 of Chap. X.) A similar effect also takes place in a magnetron if a heavy load is placed upon a long line attached to the magnetron. Assume that the Rieke diagram of Fig. 34 applies at the output-coupling device, and suppose that the impedance-changing device, instead of being close to the magnetron, is located several wavelengths away. The Rieke diagram can be transferred along the line to the impedance-changing device by rotating each point of each contour counterclockwise through an angle determined by the electrical length of the line. If, for example, the line is 10 wavelengths long at a reference frequency $f = f_3$, all points on the f_3 contour are rotated counterclockwise through 20 complete revolutions, and the

contour returns to its original position. At lower frequencies, the line length is less than 10λ, and each point of the contour is rotated through a net clockwise angle. At the higher frequencies, the contours turn through a net counterclockwise angle, since for these frequencies the line length is greater than 10λ. The power-output contours may be referred to the new location in a similar manner. The Rieke diagram at the new location is shown in Fig. 35.

In this figure line AB has been drawn as a dividing line between the contours. Actually, the contours cross this line and intersect one another, creating two sets of overlapping frequency and output-power curves for each single value of r-f load admittance. If the magnetron were a continuous-wave oscillator, either of the operating conditions might be obtained, depending upon the manner in which oscillations build up initially. If operation were to begin at point C, for example, and the r-f load changed continuously through values in the upper half of the diagram to point D, the oscillation frequencies would be below the reference frequency f_3, as specified by one set of contours. If the load admittance followed a path in the lower half of the chart to point D, the oscillation frequencies would be above f_3, corresponding to the other set of contours. Hence a different frequency and output power would result for the same r-f load admittance and input conditions.

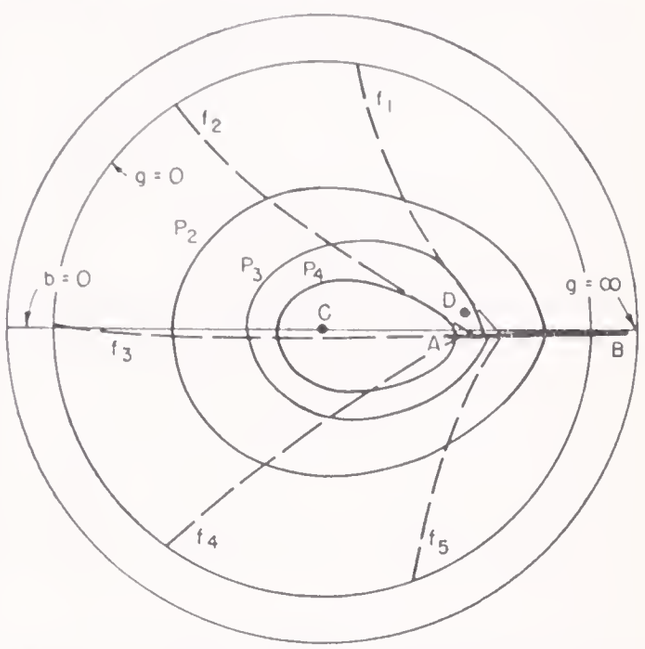


FIG. 35. Rieke diagram referred to a point of the output line 10 wavelengths distant from the reference point assumed in Fig. 34.

The above hysteresislike effect is avoided in radar magnetrons because of the pulsed condition under which they operate. If the r-f load is adjusted for operation at point D, the frequency and power will depend upon the rate at which oscillations at the two frequencies build up at the start of each pulse and upon which of the two oscillations becomes large enough to suppress the other. Thus a line such as AB exists, on one side of which the lower frequency oscillations are favored and occur in each pulse and on the other side of which the higher frequency oscillations will continually recur. In a small region near AB, an effect resembling mode skip may be expected. Oscillations in some pulses may be at the lower frequencies and in others at the higher frequencies. Because the operating point is always shifting around slightly, frequency splitting is inevitable in the region of AB. As the load is adjusted smoothly on

either side of AB, the frequency and output power also change smoothly until the load admittance crosses AB. Then a sudden jump in frequency and output power is observed.

It should be noted that line AB in Fig. 35 lies in a region of high output power, and hence the frequency-splitting effect just described is encountered only when an effort is made to extract high power from the magnetron. Although frequency *splitting* may be avoided by keeping the line section between the output-coupling and impedance-changing devices short, considerable frequency *pulling* may still be expected if the magnetron is heavily loaded.

In order to minimize the effect of load on frequency in the region of good power output, the coupling reactances and line length between resonator

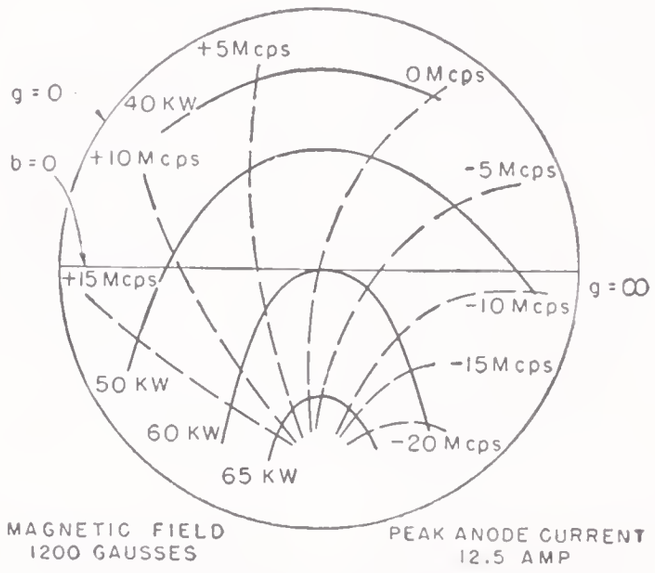


FIG. 36. Rieke diagram of a 2J32 magnetron referred to load end of quarter-wave output transformer.

and impedance-changing device are made as small as possible, and the impedance-changing device is designed to give optimum performance when the r-f line to which it is connected is terminated in a matched load. The unavoidable changes in load impedance that result from the antenna-scanning process then cause no appreciable change in operating conditions. In some magnetrons, as, for example, the Type 2J32, the impedance-changing device is an integral part of the tube and consists of a quarter-wave transformer in the r-f output connector. The Rieke diagram for this magnetron, referred to the load end of the quarter-wave transformer, has the form of Fig. 36. In other magnetrons, the quarter-wave transformer is external to the magnetron, and its position along the r-f line is under control of the user. The shape and orientation of the Rieke diagram then depend upon the impedance transformation, the location of the transformer, and the reference plane chosen for drawing the diagram.

11. Operation and Care of Magnetrons. When the magnetron is operating properly, the anode-voltage pulse is related to the magnetron-current pulse and to the r-f output-voltage pulse in the manner shown in Fig. 37. Because magnetron current becomes appreciable only when the anode voltage is near its maximum value, the current pulse is shorter and

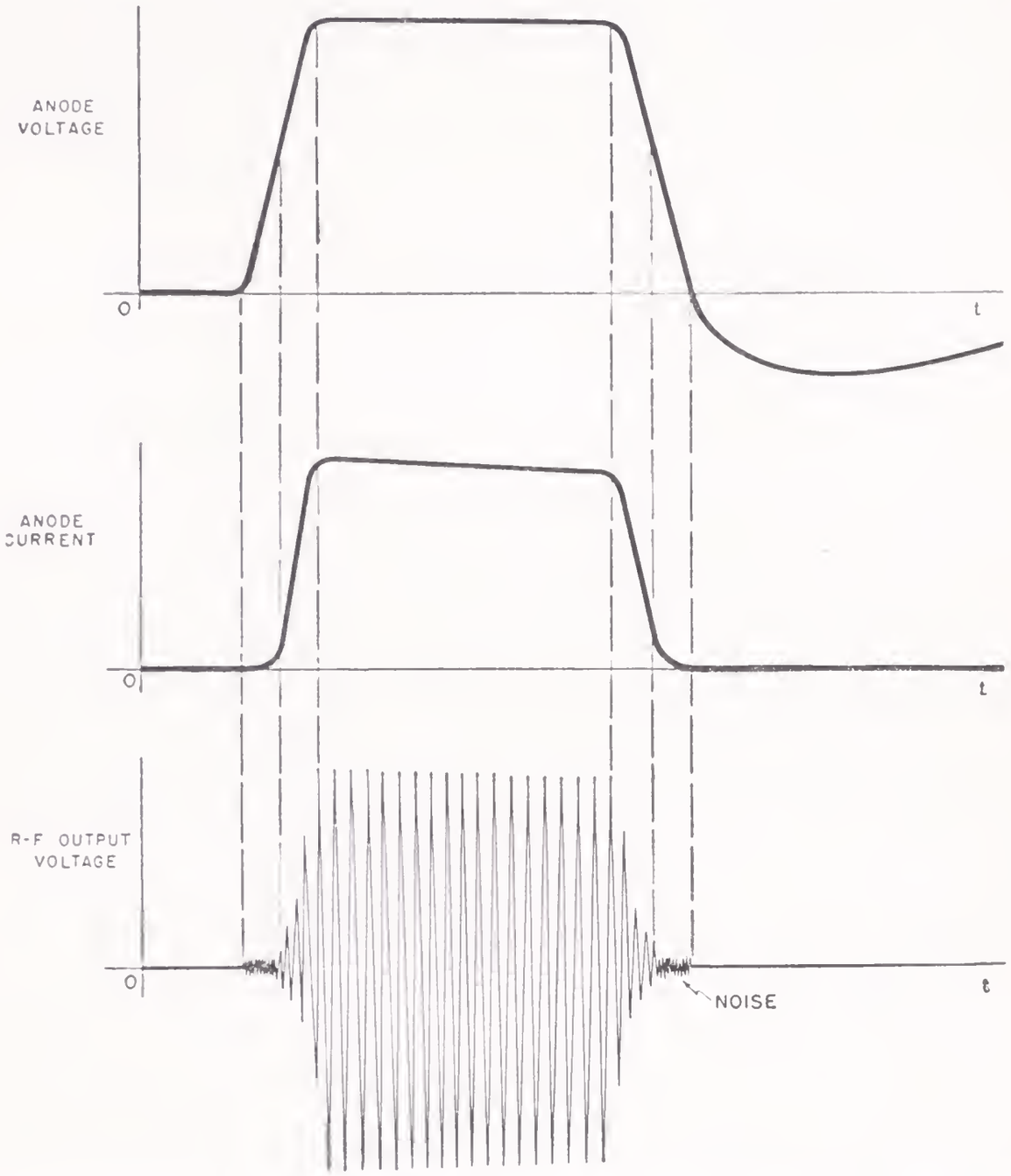


FIG. 37. Pulse shapes of magnetron anode voltage, anode current, and r-f output voltage.

has steeper sides than the voltage pulse. During the flat top of the r-f-voltage pulse, full power is produced at essentially constant frequency. During the steeply sloping sides, operation is at reduced power, and appreciable frequency changes are noted. In the short periods of low voltage at the beginning and end of the pulse, the output is mainly noise. As explained in Art. 8, the leading edge of the modulator pulse should not rise too steeply. An excessively rapid rise may cause the anode voltage to pass through the proper operating range before π-mode oscillations can

build up. Oscillations may subsequently occur in the wrong mode or fail altogether.

During operation, it is desirable to know whether or not the magnetron is oscillating consistently in the π mode. An indication of the performance may be had by observing the envelope of the r-f-voltage pulse. In order to observe the pulse envelope, a small amount of power is diverted from the r-f transmission system to an r-f envelope indicator by means of a directional coupler (Art. 7, Chap. XII). The essential components of the indicator, shown in block-diagram form in Fig. 38, are a crystal

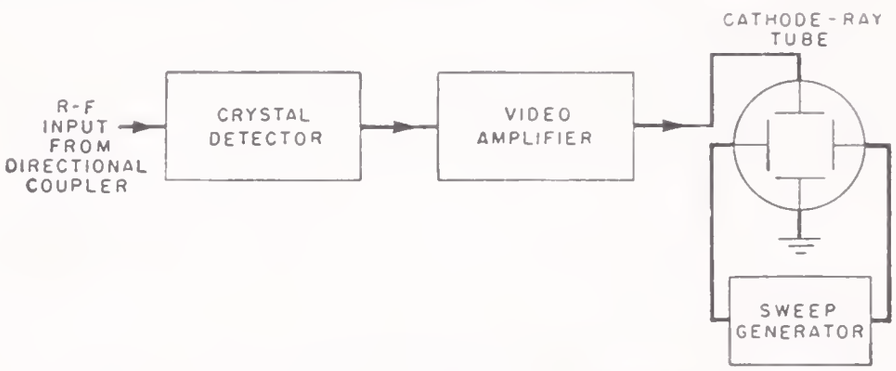


FIG. 38. Block diagram of an r-f envelope indicator.

detector, a video amplifier, and a cathode-ray tube. The r-f pulses are rectified by the detector, and the resulting video pulses after amplification are displayed on the cathode-ray-tube screen. The sweep generator is synchronized with the modulator pulses. A properly operating magnetron has an r-f pulse envelope similar to that of Fig. 39*A*. If mode skipping or frequency splitting is taking place, the envelope has the appearance of Fig. 39*B*. Mode splitting produces the envelope of Fig. 39*C*.

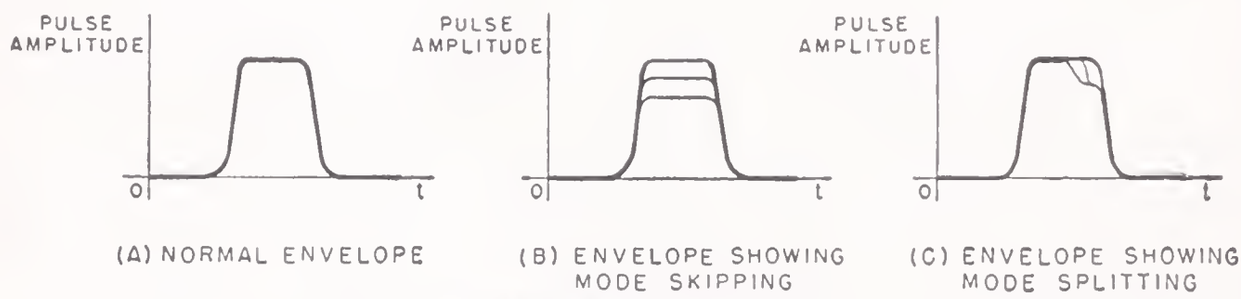


FIG. 39. Examples of r-f pulse envelopes.

Another way of evaluating magnetron performance conveniently is through observation of the frequency spectrum of the r-f pulse. For proper radar-set operation, the power in the r-f pulse should be concentrated in frequencies within the bandwidth of the receiver. Power contained in frequency components lying outside the receiver bandwidth is wasted. Examples of good and bad spectra are given in Fig. 40. Note that in the good spectrum most of the power is confined to a frequency band centered at the carrier frequency f_0. In the poor spectrum, the power is distributed over a much wider band, and consequently some of

the echo power may not pass through the receiver. One reason for a poor spectrum is that the anode-voltage pulse may not have proper amplitude to ensure operation in the π mode. Another is that the r-f line connected to the magnetron may be appreciably mismatched to the load, thus giving rise to severe frequency pulling. A poor spectrum may also result from use of a magnet of improper strength. The spectrum may be determined by means of an echo box or spectrum analyzer. Both of these instruments are described in Art. 18 of the next chapter.

A new magnetron cannot usually stand immediate application of full anode voltage. When a magnetron is being used for the first time, the pulse voltage should be increased gradually. This is especially important

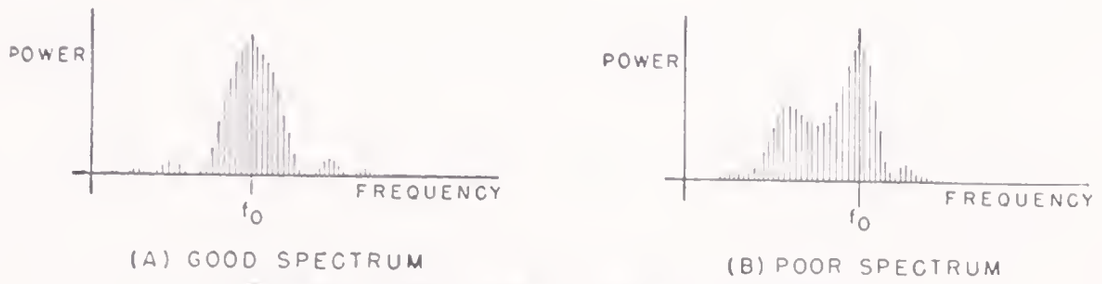


FIG. 40. Examples of magnetron spectra.

if the magnetrons are pulsed by power-amplifier modulators where the charge on the storage capacitor is much larger than is normally allowed to pass through the magnetron in one pulse. The magnets used with magnetrons are made of highly retentive alloys. In nonpackaged assemblies, the magnets are provided with keepers which should be inserted in the air space when the magnets are not in use. During operation the magnetron replaces the keeper, and under these conditions it is important not to jar the magnet or strike it with a ferromagnetic material, since such treatment may cause a marked decrease in the field strength.

Reversal of the magnetic field theoretically should not change the operation of the magnetron. Because of slight asymmetries in the magnetron assembly, however, there is usually a preferred direction for the magnetic field. Magnet-mounting instructions are normally indicated on the magnetron.

The magnetron should not be pulsed without a magnetic field. If the field is small or entirely absent, the excessive current drawn by the magnetron is likely to cause damage not only to the magnetron but also to the modulator tubes, which cannot safely carry these currents. Not only will current be excessive, but electrons, striking the anode with full velocity, may cause excessive heating. An overload relay should be provided to protect the magnetron from such damage.

CHAPTER XII

RADIO-FREQUENCY TRANSMITTING AND RECEIVING SYSTEMS

The principles of operation of two-wire transmission lines, waveguides, cavity resonators, and r-f oscillators are discussed in the preceding five chapters. The purpose of this chapter is to show how these and other r-f components are employed in the r-f transmitting and receiving system of a radar set.

The major components of the r-f system are the transmitter, the antenna, the receiver input circuit, and the interconnecting transmission lines. The short powerful bursts of radio-frequency energy generated by the transmitter are conducted over an r-f line to the antenna, where they are radiated into space. The weak echo signals from objects in space are intercepted by the antenna and directed by the r-f line to the input circuit of the receiver. In order to prevent transmitter power from entering the receiver and echo power from reaching the transmitter, a t-r device (and sometimes an anti-t-r device) must be included in the r-f line. In addition, a variety of auxiliary devices and accessories are required in order to ensure satisfactory performance of the radar set and for mechanical reasons. A description of these devices, together with a discussion of t-r and anti-t-r devices, and receiver input circuits is given in the articles that follow.

1. Radio-frequency-system Components. The type of r-f system components used in a radar set depends greatly upon the operating frequency. In sets that operate below a few hundred megacycles per second, the transmitter is a high-vacuum-triode oscillator, the antenna is usually an array of half-wave elements (dipoles), and the transmission system may be either a coaxial or parallel-wire line. At the receiver, the input stage is frequently a single-stage, radio-frequency amplifier. If an r-f amplifier is not included, echo signals are applied directly to a high-vacuum-tube type of mixer circuit.

In the microwave region, the transmitter is almost always a magnetron, and the antenna consists of a single radiating element and a metallic reflector. The r-f line may be either a coaxial line or a waveguide. Radio-frequency amplifiers are not ordinarily included in microwave receivers; the echo signals are applied directly to the mixer circuit, which employs a crystal rather than a high-vacuum tube as the mixer element.

An example of a microwave r-f transmitting and receiving system is shown in Fig. 1. The system illustrated is similar to most 10-cm systems

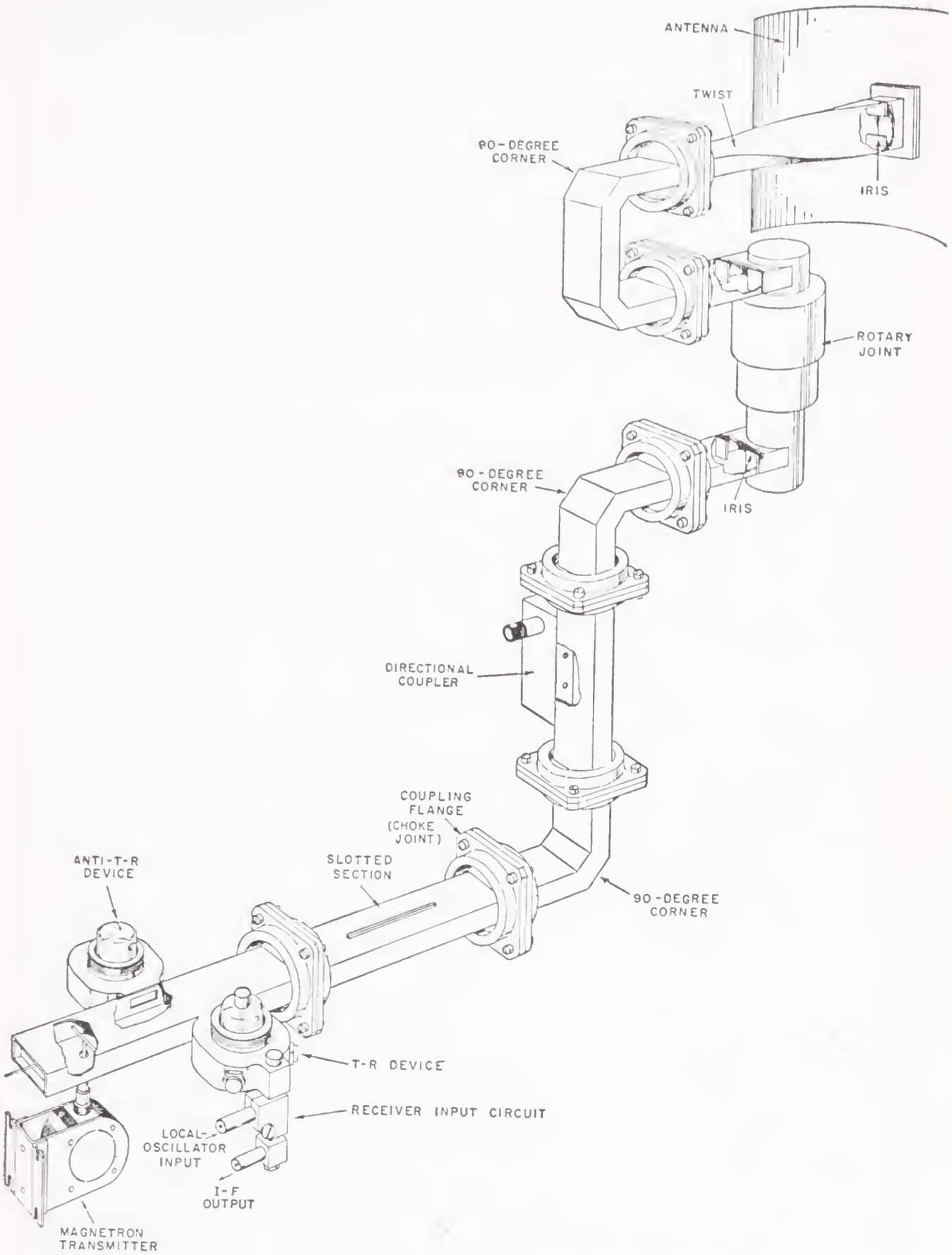


FIG. 1. Example of a 10-cm r-f transmitting and receiving system employing waveguide transmission lines.

employing waveguide transmission lines and contains many of the r-f-system components to be discussed in this chapter. Radio-frequency power is generated in pulses by the magnetron shown in the lower left-hand corner of the figure. The magnetron output power is coupled by means of a coaxial-line-to-waveguide junction into the rectangular-waveguide transmission system which directs the power to the antenna. The latter component consists of a truncated paraboloidal reflector in front of the open end of the waveguide transmission line.

A t-r device is located between the main waveguide and the receiver input circuit. This device may be thought of as a switch which disconnects the receiver from the rest of the r-f system during the time the transmitter is in operation. Isolation of the receiver is necessary during transmission in order to avoid damage to the sensitive input circuits of the receiver.

The t-r device consists of a re-entrant cavity having two points of entry—one to the main waveguide and the other to the coaxial line of the receiver input circuit. Coupling at the points of entry is provided by apertures. The re-entrant portion of the cavity is formed by two closely spaced conical-shaped electrodes which constitute a spark gap that breaks down in the presence of the high voltage of the transmitted pulse. Because breakdown of the gap prevents build-up of oscillations in the cavity, very little coupling is provided between the two points of entry, and as a result, very little transmitter power is able to reach the receiver. Echo signals from targets are able to pass through the cavity, however, because these signals are too weak to cause ionization of the t-r spark gap. Oscillations therefore build up in the t-r cavity, and echo power is coupled through the cavity to the receiver with negligible loss. Several forms of t-r devices are discussed in detail in Arts. 12 to 14.

Because power in the echo signals is extremely small (of the order of $1 \mu\mu$w), it is very important that as much of the echo power as possible be directed to the receiver and very little of it be directed toward the transmitter. In the system of Fig. 1, received signals are prevented from reaching the transmitter by the *anti-t-r device*, a re-entrant cavity aperture-coupled to the main guide at a point between the transmitter and t-r device. The anti-t-r cavity is similar in construction to the t-r cavity, except that only one point of entry is provided. The anti-t-r device, like the t-r device, serves as a switch to disconnect the transmitter from the rest of the r-f system except while pulses are being transmitted. During the reception of echo signals, the anti-t-r cavity constitutes a high-Q parallel resonant circuit at the side wall of the guide. As explained in Art. 18 of Chap. IX, the high impedance at the side wall provides essentially a short-circuit termination across the guide for waves traveling along the guide. The short circuit in the direction of the trans-

mitter at the anti-t-r location is made to appear as an open circuit at the location of the t-r device by spacing the two devices an odd-integral number of quarter wavelengths apart. Thus during the reception of echo signals, the signal path is in the direction of the receiver at the t-r device. The anti-t-r device does not interfere with the propagation of transmitter pulses through the waveguide because, during a transmitted pulse, the anti-t-r spark gap ionizes and causes the cavity to appear as a low impedance at the coupling aperture. The cavity thus has the effect of closing the coupling aperture during the transmitted pulse.

The input circuit to the receiver is a crystal mixer, the crystal being located near the lower end of a coaxial-line section aperture-coupled to the t-r cavity. Voltage from a local oscillator is coupled by means of a probe to the coaxial line, and signals at the intermediate frequency are taken from the coaxial line connected to the crystal.

The r-f system of Fig. 1 contains a variety of auxiliary components that play an important part in the over-all operation of the system. One of these components is the *rotary joint* which is located in the waveguide near the antenna. The rotary joint permits the antenna to be turned in azimuth while the rest of the set remains stationary. The antenna reflector, the upper half of the rotary joint, and the portion of waveguide above the joint are assembled as a unit and turned by means of a drive motor.

In rotary joints, mechanical contact is not usually relied upon to provide good electrical contact between fixed and movable portions of the joint. Instead, electrical connection is furnished by a radio-frequency choke in the form of a short section of shorted-end transmission line in series with each conductor of the line. Through proper choice of dimensions, the input impedance of the series-branch line can be made extremely low, and good electrical connection between main-line conductors thus provided. Joints which include a transmission-line type of r-f choke are commonly referred to as *choke joints*.

Choke joints are also used frequently to couple sections of waveguide together, as indicated in Fig. 1. Use of a choke joint ensures good electrical connection between adjacent guide sections and avoids the need for accurate alignment of parts being joined.

Because the waveguide transmission system of Fig. 1, like most transmission systems, follows an irregular path between transmitter and antenna, several transitional sections of guide in the form of *corners* and *bends* are required. A 90-deg *twist* is also placed in the guide near the antenna so that the antenna reflector may be illuminated with waves which are horizontally polarized. The dimensions of the various transitional sections are chosen so that as little reflection as possible is set up at the ends of the section.

The irises shown in Fig. 1 are impedance-changing devices used to ensure that the major portion of the waveguide system operates with essentially unity standing-wave ratio during the transmission of pulses. At the transmitter end of the waveguide, the movable plunger and waveguide-to-coaxial-line junction provide the proper load impedance for the magnetron output line.

The performance of the over-all r-f system of Fig. 1 may be evaluated from measurements made with the aid of the *slotted guide* section and the *directional coupler*. The slotted guide makes possible the measurement of the standing-wave ratio in the guide, and the directional coupler makes available a small amount of r-f power for determining such quantities as the transmitter frequency, the frequency spectrum of the transmitted pulse, and the sensitivity of the receiver. Details of these measurements and of the equipment involved are discussed in the last article of this chapter.

2. Choke Joints. As explained in the preceding article, choke joints enable two line sections to be joined electrically despite lack of good mechanical contact between sections. These joints depend for their operation upon shorted-end transmission lines in series with the conductors being joined. Constructional details of a coaxial-line type of choke joint are shown in Fig. 2*A*. Auxiliary line 1 is a quarter-wave shorted-end section of coaxial line formed by the outside surface of the outer conductor of the main line and inside surface of the surrounding metallic cylinder. The input impedance Z_1 of this line, in series with the resistance at point A (contact resistance or radiation resistance), constitutes the load impedance for auxiliary line 2, which is also a quarter wave in length. Since Z_1 is nearly infinite, the input impedance Z_2 of line 2 is essentially zero, and negligible impedance is introduced at point B of the main line. The inner conductors are joined electrically in a similar manner, dimensions being chosen so that a low impedance Z_3 appears at point C.

An important feature of the choke joint of Fig. 2*A* is that any resistance at point A which may result from poor mechanical contact or from radiation at an open gap is in series with the high input impedance of line 1. Thus current flow in the resistance and consequently the power loss are kept extremely low.

In order that the impedance Z_2 at the junction of the main-line outer conductors may be kept low over a band of frequencies, the characteristic resistance of line 1 is usually made somewhat greater than that of line 2. Making the value of R_c for line 1 as large as feasible yields a high value of Z_1 over the frequency band of interest; making R_c for line 2 as low as possible ensures a low value of Z_2 at point B over the same frequency band.

Further improvement in the broad-band operation of the over-all joint may be obtained by spacing the open ends of the branch lines (points B and C) a quarter wavelength apart. At the frequency for which the branch-line sections are each a quarter wave, Z_2 and Z_3 are both resistive, but at slightly different frequencies, both contain reactive components. By separating points B and C a quarter wavelength, the effect of the reactive components on the standing-wave ratio beyond the joint can be made to cancel approximately.

Alternate forms of choke-joint construction are shown in Fig. 2*B*. In the upper diagram, one portion of the branch line in series with the outer

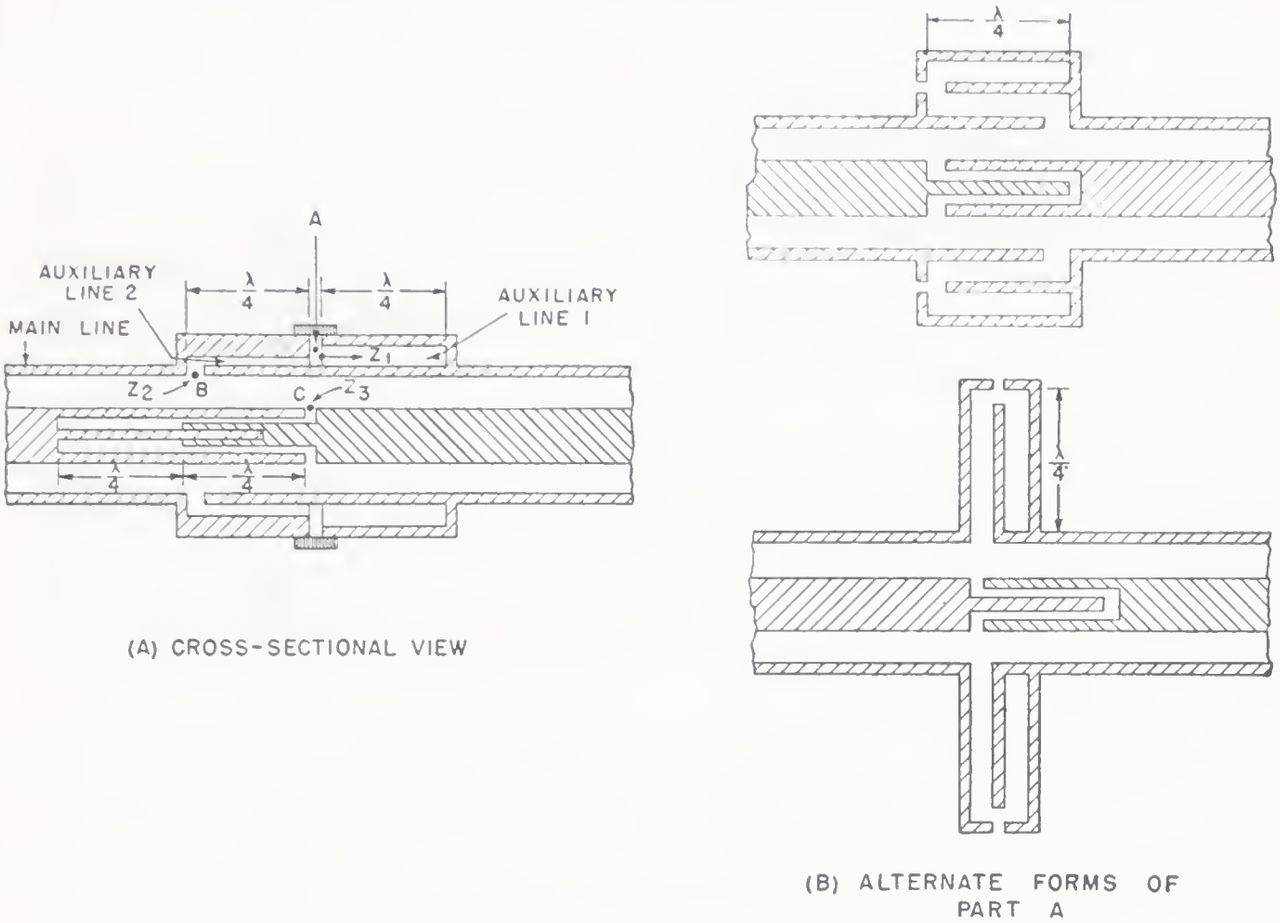


FIG. 2. Coaxial-line choke joints.

conductors is folded back upon the other, and in the lower diagram, the branch lines are extended radially outward from the main line. In both diagrams the branch lines in series with the inner conductors are made approximately a quarter wave in length and are terminated essentially in an open circuit rather than a short circuit.

Choke joints are also used to couple magnetrons having coaxial-line output terminals to the main coaxial line. Rigid mechanical connection to the magnetron output line must be avoided because of the danger of breaking the fragile metal-to-glass seal. An example of a magnetron coupling device is given in Fig. 3.

Choke joints between two sections of waveguide are shown in Fig. 4.

In Fig. 4*A*, line 1 is a coaxial line formed by cutting a slot in a flange surrounding the guide, and line 2 is a radial transmission line formed by the faces of the two coupling flanges. If a $TM_{0,1}$ wave is being propagated in the circular guide, the electric-field pattern in the guide is axially symmetrical, and transverse-electromagnetic-wave patterns are set up in the series-branch lines constituting the choke joint. If each branch line is made a quarter wavelength as measured in space, negligible impedance is introduced in series with the two guide sections being joined. Broadband operation is obtained in the same manner as with coaxial lines, that is, by making the characteristic resistance of line 1 high compared with that of line 2. In practice it is found that a gap of several millimeters can be allowed between faces of the two flanges without introducing serious reflections in the guide.

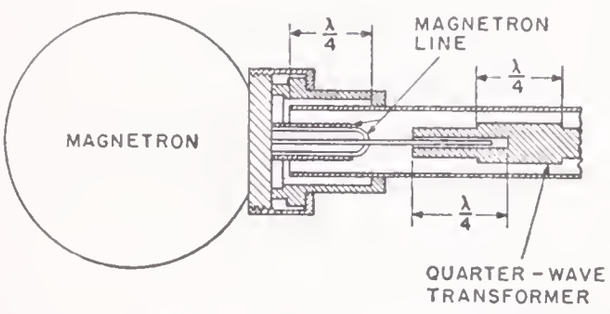


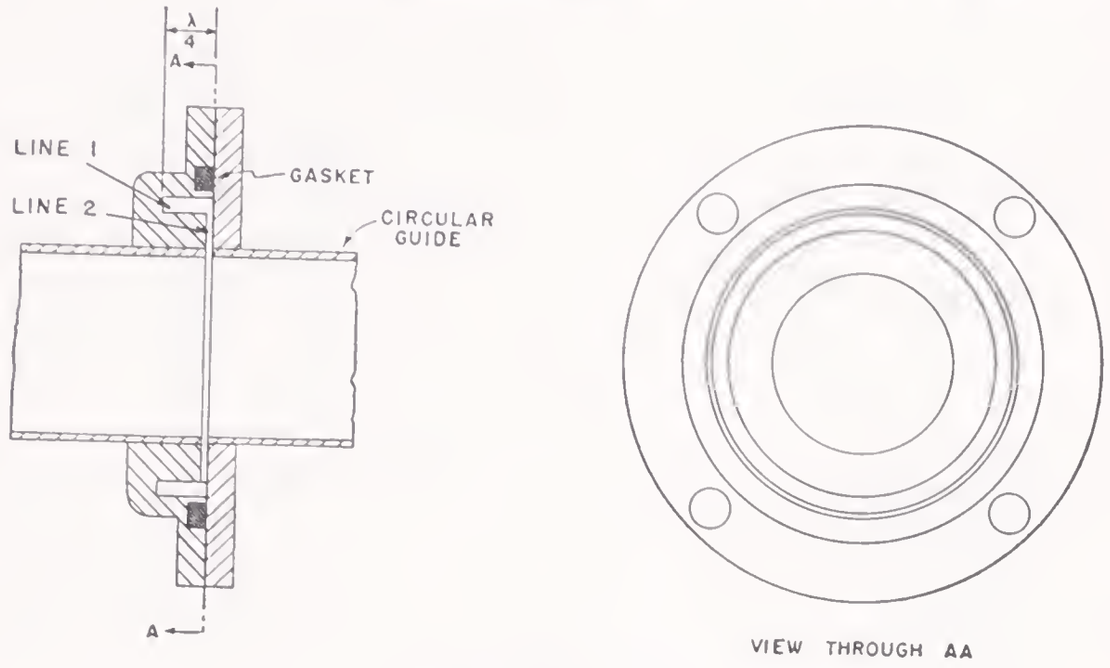
FIG. 3. Magnetron output-coupling device.

If the guide has a rectangular cross section, as in Fig. 4*B*, and $TE_{1,0}$ waves are being propagated, higher order waves rather than principal waves are set up in the series-branch lines. These higher order waves result because of the asymmetric nature of a $TE_{1,0}$ wave about the longitudinal axis of the guide. Each branch line should be made approximately a quarter wave for the wave type excited. Because the actual field pattern in the branch lines may consist of two or more higher order waves, it is usually necessary to determine the lengths experimentally at the operating wavelength.

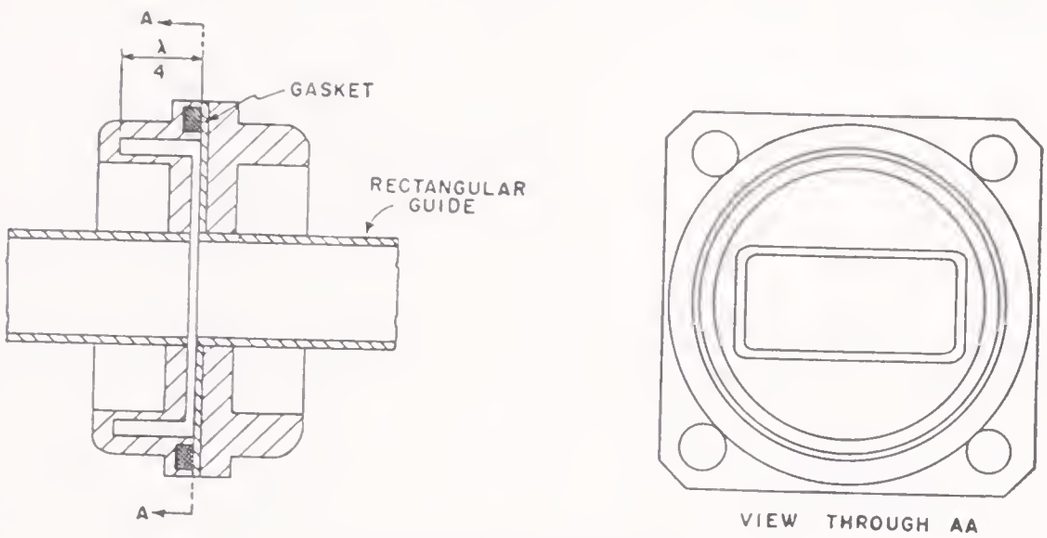
3. Rotary Joints. Rotary joints enable portions of the r-f transmission line to be rotated with respect to one another. Rotation of transmission-line sections is necessary whenever the antenna must be pointed in different directions in space. For proper operation of the antenna, it is necessary that the field patterns of waves propagated across the joint be symmetrical about the axis of rotation. Otherwise, the polarization of the waves in the movable portion of the line would change as the line turns and, as a result, cause undesirable radiation patterns from the antenna.

Rotary joints for coaxial-line r-f systems are easily constructed, since the transverse-electromagnetic-wave pattern of a coaxial line is axially symmetrical. The simplest form of coaxial-line rotary joint utilizes fingerlike rubbing contacts on the inner and outer conductors. Such a joint is satisfactory for low-power r-f systems in which the turning speed of the rotating components is low, but it has the disadvantage that continual wear at the rubbing surfaces ultimately results in poor electrical contact and hence excessive power loss.

A more reliable type of joint employs the choke-coupling principle described in the preceding article. The important electrical and mechanical features of such a joint are illustrated in Fig. 5. The upper line is arranged to turn while the lower line is held stationary. Half-wave chokes serve to connect corresponding conductors of the lines together



(A) CIRCULAR GUIDE



(B) RECTANGULAR GUIDE

FIG. 4. Waveguide choke joints.

electrically. Bearings are provided at the rubbing surfaces. The pressurizing seal located near the outer surface of the joint serves to make the joint airtight and allows the line to be filled with air at a pressure above atmospheric pressure. It is common practice to pressurize radar lines in this manner in order to lessen the danger of breakdown of the dielectric between conductors due to moisture in the line or low air pressure at high altitudes.

Construction of a rotary joint is more difficult for a waveguide transmission system than for a coaxial-line system because the guides are usu-

ally rectangular, and energy is propagated in an axially unsymmetrical ($TE_{1,0}$) wave. In order to obtain an axially symmetrical field pattern in the vicinity of the joint, a short section of circular guide carrying a $TM_{0,1}$ wave is usually employed.

In order to excite a $TM_{0,1}$ wave in a circular guide from a dominant wave in a rectangular guide, advantage is taken of the fact that the magnetic- and electric-field patterns in planes parallel to the top wall of a rectangular guide propagating a dominant wave bear a resemblance to the corresponding $TM_{0,1}$ patterns in transverse planes of a circular guide. (Compare the rectangular-guide $TE_{1,0}$ pattern, Fig. 13, p. 573, with the circular-guide $TM_{0,1}$ pattern, Fig. 33, p. 609.) It is possible, therefore, to excite the $TM_{0,1}$ waves simply by inserting the circular section into the top or bottom wall of the rectangular guide. A rotary joint based upon this method of construction is shown in Fig. 6*A*. The two portions of circular guide are joined together electrically by means of a choke joint, and irises are inserted in the regular guide to cancel any reflections that may occur at the junctions of guide sections.

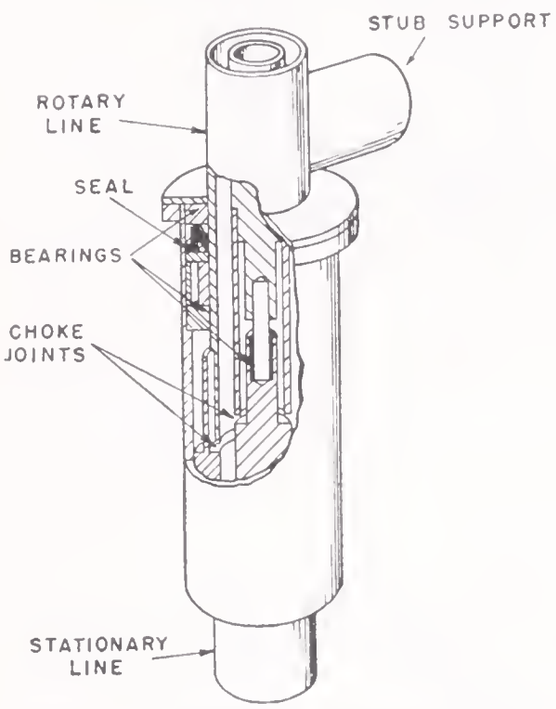


FIG. 5. Coaxial-line rotary joint.

Because dominant ($TE_{1,1}$) waves, as well as $TM_{0,1}$ waves, may be propagated in the circular-guide section, care must be taken to avoid exciting $TE_{1,1}$ waves of appreciable amplitude. Otherwise a field pattern would result which would be unsymmetrical about the axis of the guide. In practice it is found possible to concentrate most of the transmitted power in $TM_{0,1}$ waves through careful choice of the diameter of the circular guide. The circular-guide diameter is most readily determined experimentally in relation to the size of rectangular guide being used and the operating frequency.

An additional precaution against $TE_{1,1}$ waves may be taken by extending the circular-guide section beyond the rectangular guides, as shown in Fig. 6*B*. The distance from the center of the rectangular guide to the shorted end of the extended section is made an odd multiple of quarter waves (usually $\frac{3}{4}\ \lambda$) for $TE_{1,1}$ waves. It is possible, through proper choice of the diameter of the extended section, to make this distance correspond to $\frac{1}{2}\ \lambda$ for $TM_{0,1}$ waves. Thus the input impedance of each extension is essentially infinite for $TE_{1,1}$ waves and zero for $TM_{0,1}$ waves. With reference to $TM_{0,1}$ waves, therefore, the joint is equivalent to that

of Fig. 6A. Since the infinite impedance of the extended section and the remaining circular guide may be regarded as being in series across the rectangular guide, $TE_{1,1}$ waves are not transmitted through the joint.

An alternative method of launching $TM_{0,1}$ waves in the circular guide makes use of a probe which extends across the rectangular guide and

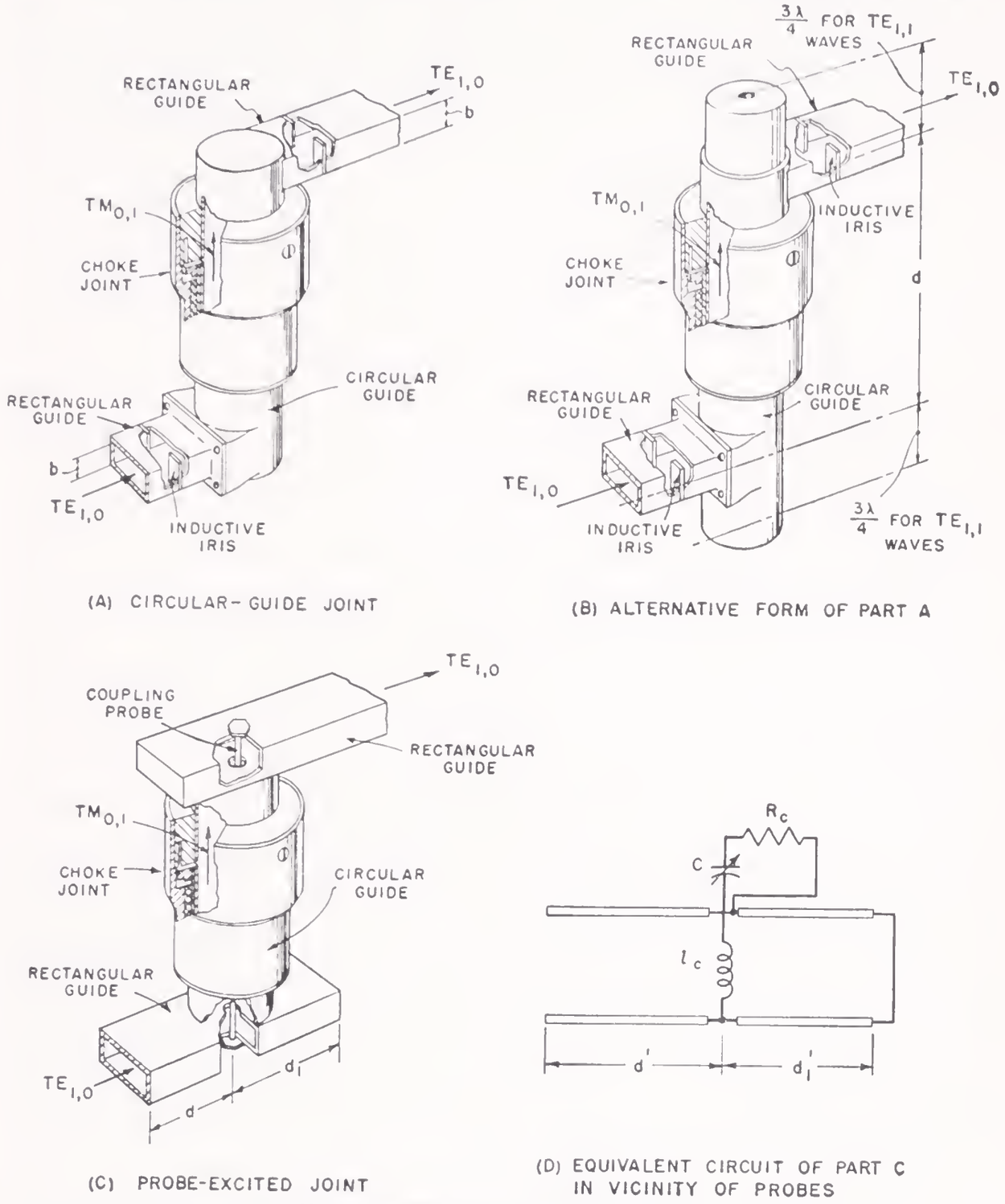


FIG. 6. Waveguide rotary joints.

penetrates into the circular guide, as shown in Fig. 6C. The probe and surrounding cylinder may be thought of as a short section of coaxial line, the principal (transverse-electromagnetic) wave of which gives rise to a $TM_{0,1}$ wave in the region beyond the probe. (Compare the field patterns of transverse-electromagnetic waves in coaxial lines with $TM_{0,1}$ waves in circular guides.)

The junction between circular and rectangular guides is similar to the waveguide-to-coaxial-line junction illustrated in Fig. 52*A* of Chap. IX, except that only one tuning adjustment—the depth of penetration of the probe into the circular guide—is provided. The equivalent circuit of the junction (Fig. 6*D*) is therefore analogous to the equivalent circuit of the waveguide-to-coaxial-line junction shown in Fig. 52*B*, Chap. IX. Distance d_1 from the center of the probe to the end wall of the rectangular guide is chosen so that it is possible to match the rectangular guide for unity standing-wave ratio by means of the single probe adjustment.

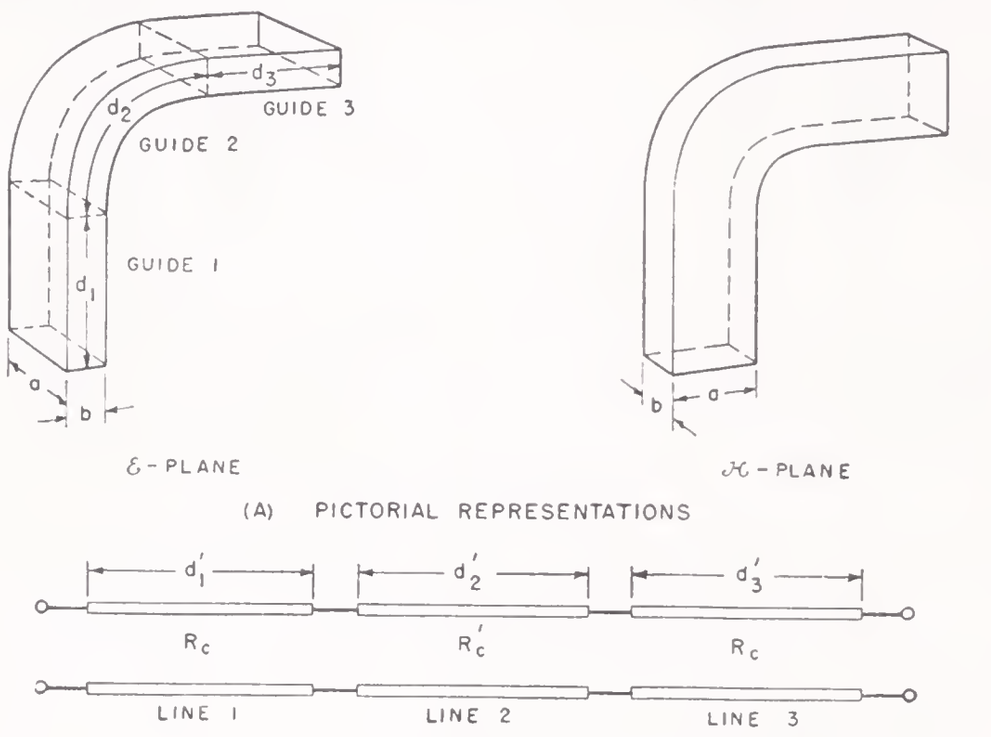


FIG. 7. Circular bends in rectangular waveguides.

4. Waveguide Bends, Corners, and Twists. The path of a waveguide transmission system may be changed by inserting a corner section of guide in the system or by bending the guide smoothly through the desired angle. Corners are usually employed where an abrupt change in direction is required. If the change can be made gradually, a smooth bend is often preferable to a corner because of the ease with which the bend can be formed. Low-reflection bends of large radius can be made simply by curving a straight piece of guide.

Two types of circular bends for rectangular waveguides are illustrated in Fig. 7*A*. In the E-plane bend, the guide is curved in such a way that the electric-field vector of the propagating $TE_{1,0}$ wave remains in the same plane. In the H-plane bend, the guide is bent so that the magnetic-field vector remains in the same plane throughout the bend.

It is desirable to make bends with as large a radius of curvature as possible in order that appreciable reflections at the ends of the curved sec-

tion may be avoided. Sharp bends tend to give rise to higher order waves at the junctions of the straight and curved sections because of the abrupt changes in direction. These higher order waves, although not propagated, have the effect of introducing a reactive component into the wave-guide impedance at the points of discontinuity.

When the bending radius is large, the dominant wave in the curved guide has field patterns in transverse planes which are similar to the corresponding patterns of the dominant wave in a straight guide. However, in the curved guide, the wave impedance and phase constant are slightly different from the corresponding straight-guide quantities.[1] The change of wave impedance may be accounted for in the two-wire equivalent circuit of the bend by an appropriate change in the characteristic resistance of the line section representing the curved section of guide. Thus in Fig. 7*B*, the middle line, representing the curved portion of the bends in Fig. 7*A*, has a characteristic resistance R_c' equal to the wave impedance of the dominant wave in the curved-guide section. The electrical length of the curved-guide section and the equivalent transmission-line length are also made equal and are measured in terms of the dominant-wavelength in the curved section. Figure 7*B* indicates that if guide 3 is terminated in a matched load, guide 1 may be similarly terminated if the curved section is made an integral number of half curved-guide wavelengths. The curved-guide section is then equivalent to a 1:1 transformer connecting the straight sections.

If the bending radius is short, series and shunt reactive elements must be included in the two-wire equivalent circuits at the input and output terminals of the middle line in order to account for the higher order waves set up at the ends of the curved guide. The presence of these elements makes complete elimination of reflections impossible, but even in sharp bends the straight and curved sections merge so smoothly that reflections are usually small.

Right-angle corner sections of the $\mathcal{E}$-plane and $\mathcal{H}$-plane types are illustrated in Figs. 8*A* and *B*. These sections may be thought of as being composed of three straight sections of guide with ends beveled so as to give the desired change in direction. Because the directional changes occur more abruptly in corner sections than in bends, reflections at the points of discontinuity are relatively greater in the corner sections.

A two-wire equivalent circuit of the $\mathcal{E}$-plane and $\mathcal{H}$-plane corners is drawn in Fig. 8*C*. Observe that the characteristic resistances of the transmission-line equivalents of the straight-guide sections are all alike and are equal to the wave impedance of the dominant wave in the

[1] For the derivation of the wave impedance and phase velocity of the dominant wave in a curved rectangular guide, see S. A. SCHELKUNOFF, *Electromagnetic Waves*, 324–330 (D. Van Nostrand Company, Inc., New York, 1943).

straight-guide sections. The electrical length of each line is equal to the mean electrical length of the corresponding guide section.

The equivalent circuit of Fig. 8*C* is rather general in that it does not specify the network of reactive elements required to account for the higher order waves generated at the junctions. The length of line 2 which yields minimum reflections cannot, therefore, be specified. An approximate value may be obtained for the E-plane corner if it is assumed that the networks are single shunt capacitive elements of equal value and small magnitude. Such an assumption is reasonably valid provided the change in angular direction at each junction does not exceed 45 deg. If single shunt capacitive elements are assumed at each end of line 2, it may be shown that making line 2 an odd number of quarter waves yields essentially a flat line for line 1 when line 3 is terminated in a matched load.

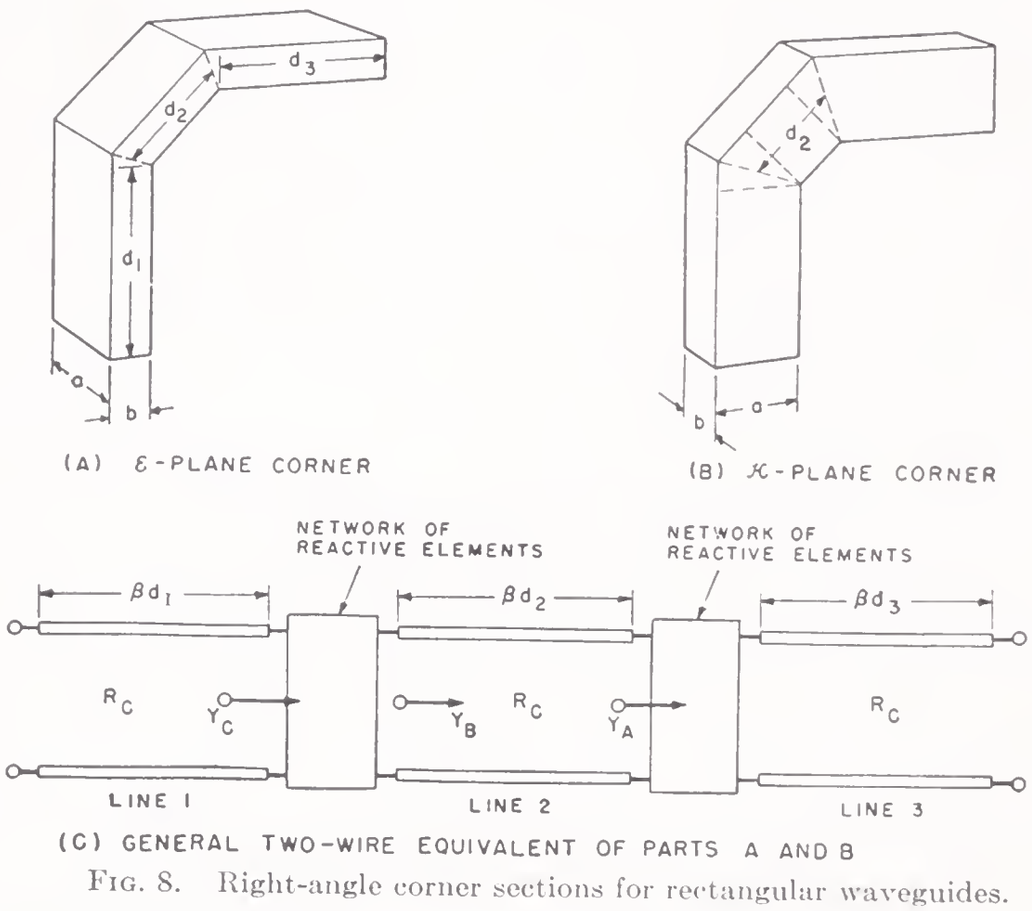


FIG. 8. Right-angle corner sections for rectangular waveguides.

Consider, for example, that each network in Fig. 8*C* presents a per-unit shunt susceptance b_c across input and output terminals of line 2 and that line 3 is terminated in a matched load. The per-unit admittance y_A at the output terminals of line 2 is then $y_A = 1 + jb_c$. If line 2 is a quarter wave, its input admittance y_B is very nearly the conjugate of y_A when b_c is much less than 1; that is, $y_B \approx 1 - jb_c$. The shunt-element susceptance b_c at the input terminals of line 2 is therefore canceled by the susceptance of y_B, and line 1 is terminated in a matched load.

If the angular change in direction of the E-plane corner exceeds 45 deg, series as well as shunt elements are needed to represent junction effects

accurately, but the proper length for line 2 is still very nearly a quarter wavelength. In general both series and shunt reactive elements are always required to represent junction effects in 3C-plane corners, and reflections are minimum if line 2 in the equivalent circuit is made slightly greater than a quarter wave.[1]

Twists are sometimes inserted in rectangular-guide r-f systems in order to change the direction of polarization of the waves emitted from the end of the guide. Twists should be made gradually over several wavelengths and should be approximately an integral number of half waves in length in order that slight changes in wave impedance and phase constant in the twisted section will cancel at the ends.

5. T Junctions. Often in an r-f transmission system three sections of line or guide are joined together at right angles to one another. Such a junction is referred to as a T junction. Two basic forms of equivalent circuits of T junctions are shown in Fig. 9. These circuits correspond to parallel and series connection of lines and are accurate representations of a T junction provided the transverse dimensions of the component lines are negligible compared with the operating wavelength. Nearly all

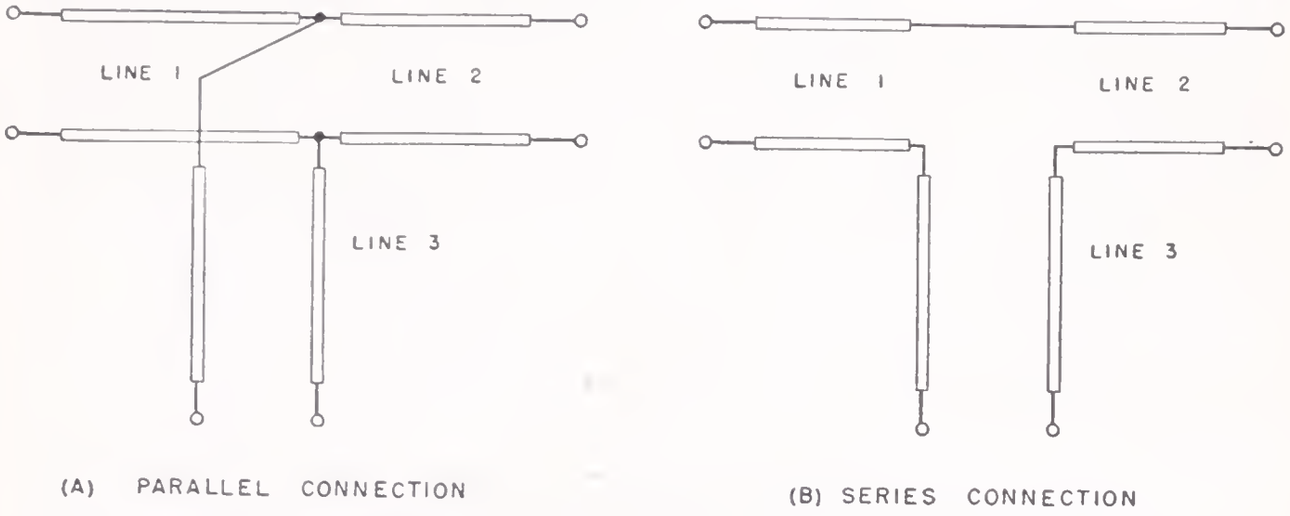


FIG. 9. Basic forms of equivalent circuits of a T junction.

parallel-wire lines meet this requirement, but very often transverse dimensions of coaxial lines are an appreciable portion of a wavelength. The circuits of Fig. 9 are then only approximate representations, additional elements being necessary to account for discontinuity effects at the junction of lines. A more accurate equivalent circuit of a coaxial-line T junction at high frequencies includes a coupling inductance in series with the side-arm line of Fig. 9*A* (see Art. 13, Chap. IX).

Of interest in connection with rigid coaxial lines are the T junctions formed by the intersection of the main line and the shorted-end lines (stubs) that provide support for the main-line inner conductor. In one simple form of stub support, the main line is straight and the stub is

[1] See G. L. RAGAN, *Microwave Transmission Circuits*, 203–207 (McGraw-Hill Book Company, Inc., New York, 1948).

at right angles to the main line, as shown in Fig. 10*A*. All lines have the same characteristic resistance. The equivalent circuit of the support, drawn in accordance with the principles of Art. 13 of Chap. IX, is given in Fig. 10*B*. It is evident from this circuit that stub length d_a should be a quarter wave at the operating frequency in order that reflections at the junction may be avoided. Coupling inductance l_c is then in series with the extremely high impedance of the stub, and the combination appears essentially as an open circuit across the main line.

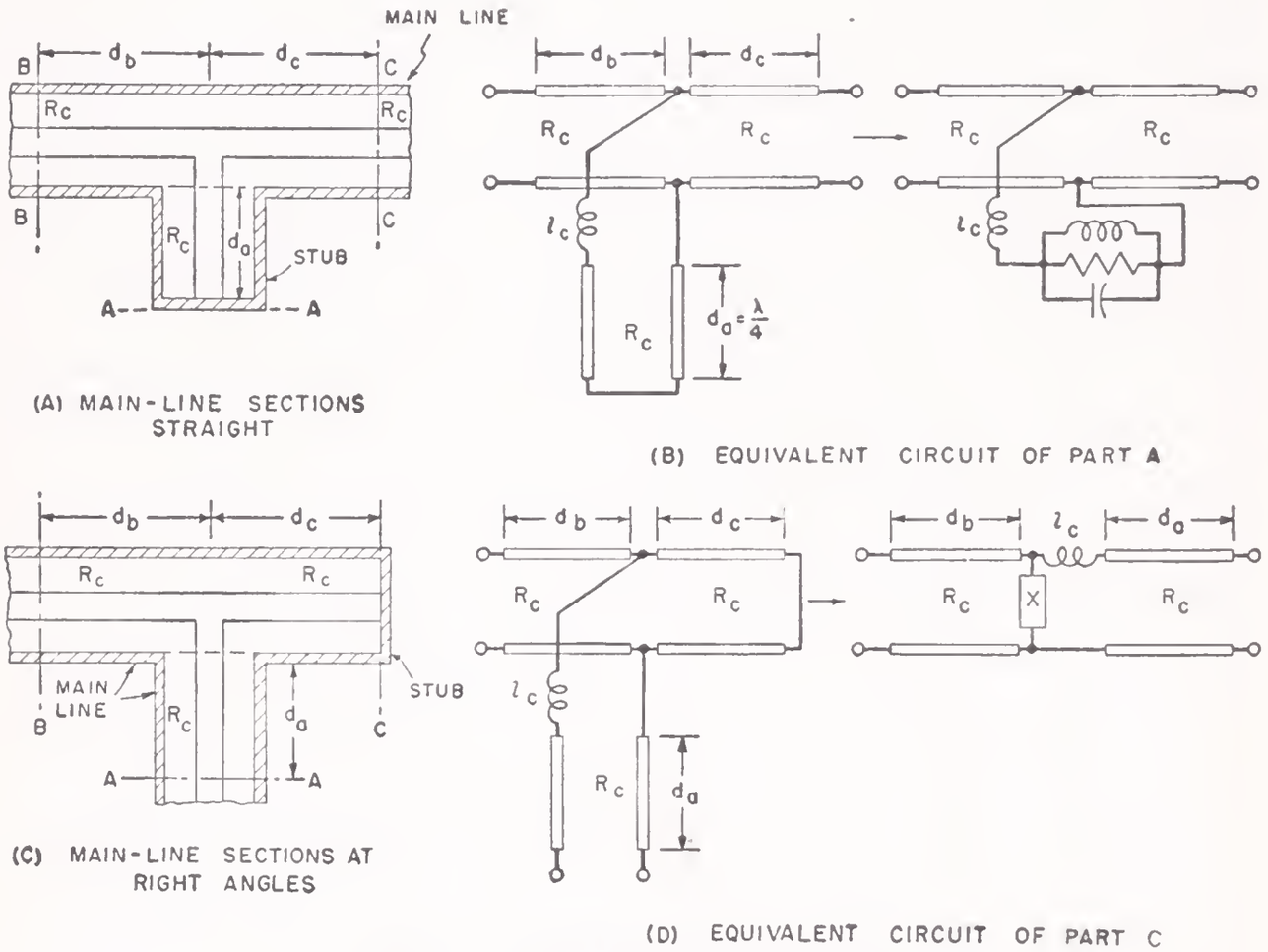


FIG. 10. Coaxial-line stub supports.

If a change in direction of the main line is desired, the positions of stub and one main-line section may be interchanged, as shown in Fig. 10*C*. With the stub in the new position, however, the main-line sections cannot be completely matched no matter what length is chosen for the stub support. Consider the equivalent circuit of Fig. 10*D*. Because one main-line section is now the side arm of the T junction, the coupling inductance l_c appears in series with the main line, and the stub reactance X is in shunt with the series combination of l_c and the input impedance of the side-arm line. If the main line is terminated in a matched load at point A, no adjustment of the stub will result in a matched load for the remaining main-line section.

A disadvantage of the simple stub support of Fig. 10*A* is that it is sensitive to small changes in frequency, a perfect match of main-line

sections being obtainable only at the frequency for which the stub is quarter-wave resonant. A stub support that yields essentially a flat line over a broad band of frequencies may be constructed by enlarging the inner conductors of the component lines in the vicinity of the T junction (see Fig. 11*A*). The equivalent circuit of the broad-band stub (Fig. 11*B*) is similar to the equivalent circuit of the simple T junction except that the component line sections have a slightly lower characteristic resistance than that of the main line. Usually lines 2 and 3 have equal values of R_c and line 1 a slightly different value.

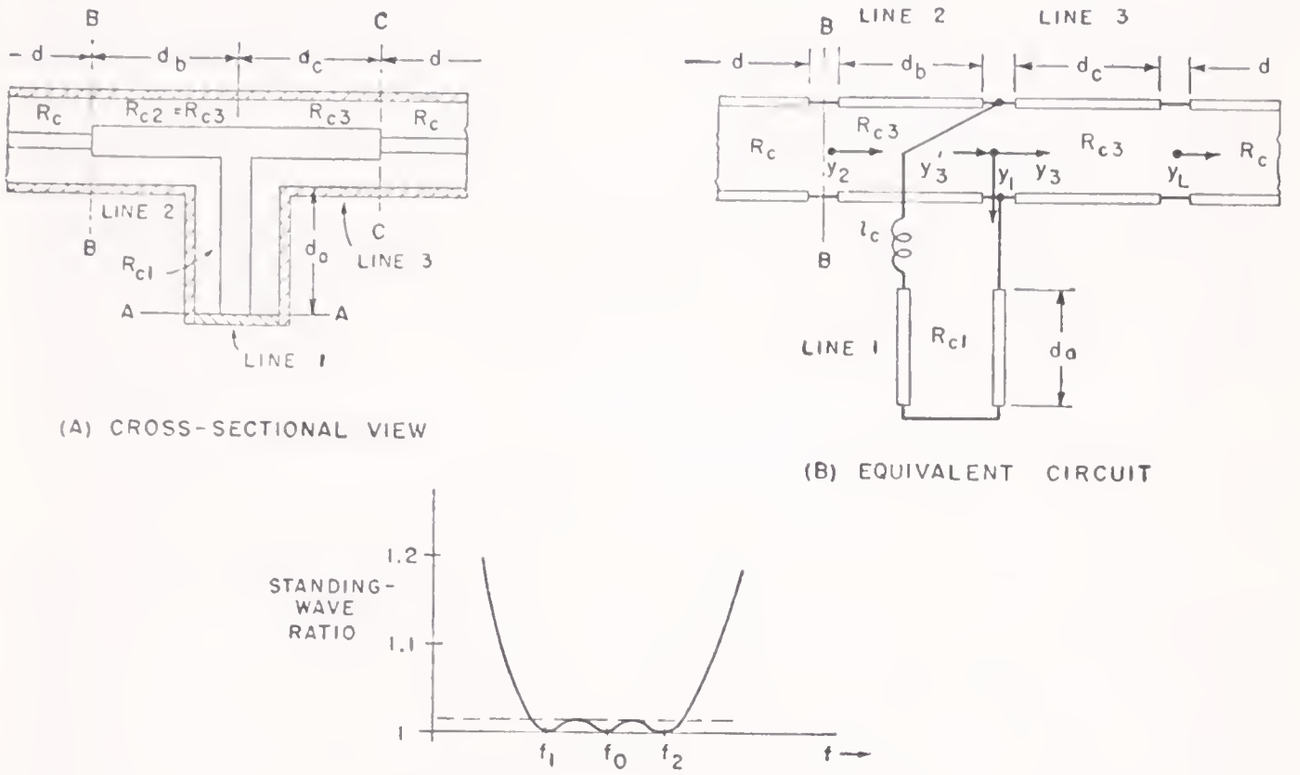


FIG. 11. Broad-band coaxial-line stub support.

The broad-band property of the stub is the result of the reflections set up in the vicinity of the stub. Through proper choice of diameter and length of the enlarged conductors, reflections are made to cancel completely at three frequencies in a band, and nearly completely at the remaining frequencies within the band. A typical curve of standing-wave ratio introduced in the main line by a single stub is plotted in Fig. 11*C* as a function of frequency.

The frequency characteristic of the stub may be explained in more detail with the aid of the equivalent circuit and the Smith chart. If the stub is terminated in a resistance R_c at point C and distances d_a, d_b, and d_c are each made equal to a quarter wave at frequency f_0, a flat line to the left of point B is obtained because line 1 and l_c appear essentially as an open circuit at the junction of lines, and lines 2 and 3 together are equiv-

alent to a one-to-one (half-wave) transformer. On the Smith chart (Fig. 12*A*), $y_L = G_c/G_{c3}$, the per-unit load admittance terminating line 3, and y_2, the per-unit input admittance of line 2, are therefore located at a common point.

At frequencies below f_0, lines 1, 2, and 3 are each less than a quarter wave, but a particular frequency f_1 below f_0 may be located at which $y_2 = y_L$. At such a frequency d_c is less than a quarter wave, and the input admittance y_3 of line 3 is located at the point marked y_3 in Fig. 12*B*. Adding y_1, the inductive susceptance of line 1 to y_3 yields y_3', the load admittance of line 2, and rotating point y_3' in the chart clockwise through a distance $d_b = d_c$ yields the point $y_2 = y_L$.

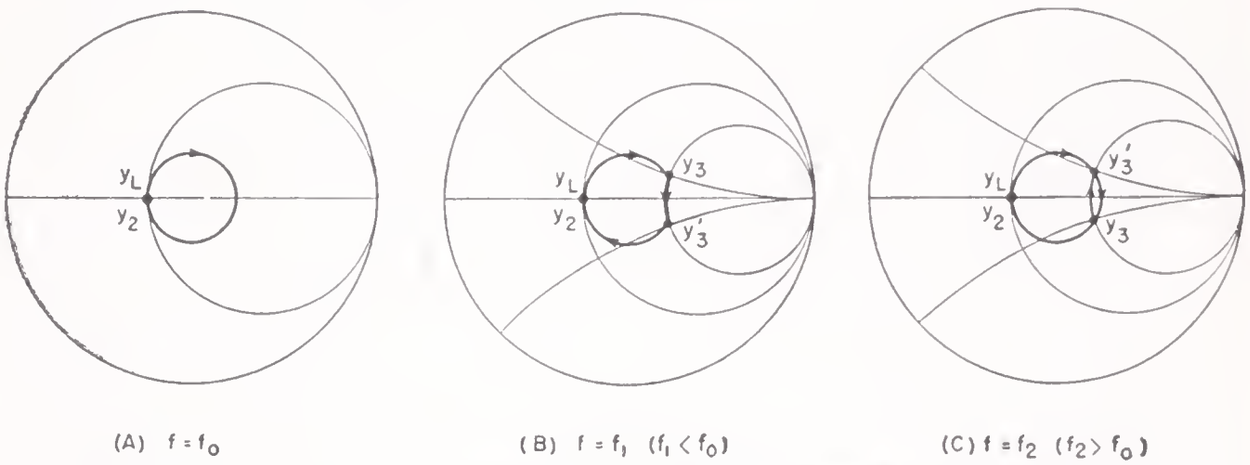


FIG. 12. Admittance relations for a broad-band coaxial-line stub support.

Observe that a match of main-line admittances is possible at frequency f_1 because the susceptance in the direction of the stub is opposite in sign to the susceptance component of y_3 and of such magnitude as to make the complex number y_3' the conjugate of the complex number y_3. The stub is thus able to compensate exactly for the fact that $d_c + d_b$ is less than a half wave.

A frequency f_2 above f_0 may also be found at which a conjugate relation exists between y_3 and y_3'. Admittances at the ends of the various line sections are indicated on the Smith chart of Fig. 12*C*. At this frequency, the susceptance in the direction of the stub exactly compensates for the fact that $d_c + d_b$ is greater than a half wave.

In general, the greater the change in main-line characteristic resistance at the stub, the more widely separated are the frequencies at which a perfect match is obtained. As the spacing of the critical frequencies is increased, however, a relatively greater mismatch occurs at the intermediate frequencies. Broad-band stubs are commonly designed to operate over a range of ± 10 per cent of the mid-band frequency. The standing-wave ratio introduced in the main line by the stub may be maintained less than 1.02 over the band (see Fig. 11*C*).

A relation between the characteristic resistances of the component lines of the stub support and the frequencies at which a perfect match of main-

line sections is obtained may be derived rather simply if it is assumed that l_c is negligible and that $d_a = d_b = d_c = d$. The load admittance terminating line 3 in Fig. 11*B* is $Y_L = 1/R_c$, and the input admittance of the same line, which may be determined by manipulation of Eq. 41, p. 496 of Chap. VII, is

$$Y_3 = G_{c3} \frac{Y_L + jG_{c3} \tan 2\pi d/\lambda}{G_{c3} + jY_L \tan 2\pi d/\lambda} = \frac{1}{R_{c3}} \frac{1/R_c + j1/R_{c3} \tan 2\pi d/\lambda}{1/R_{c3} + j1/R_c \tan 2\pi d/\lambda} \tag{1}$$

where $G_{c3} = 1/R_{c3}$ is the characteristic conductance of line 3 and $2\pi d/\lambda$ is the electrical length of the component lines at wavelength λ. If l_c is neglected, the input admittance of the stub is

$$Y_1 = -j \frac{1}{R_{c1}} \cot \frac{2\pi d}{\lambda}$$

As explained above, the requirement for a perfect match of lines is that Y_1 must be twice the negative value of the imaginary component of Y_3. Thus if Eq. 1 is rationalized and the imaginary part equated to $-\frac{1}{2}Y_1$, the result is

$$\frac{1}{R_{c3}} \frac{1 - (R_{c3}/R_c)^2}{\cot 2\pi d/\lambda + (R_{c3}/R_c)^2 \tan 2\pi d/\lambda} = \frac{1}{2} \frac{1}{R_{c1}} \cot \frac{2\pi d}{\lambda} \tag{2}$$

A rearrangement of terms in Eq. 2 leads to the following expression:

$$\cot^2 \frac{2\pi d}{\lambda} = 2 \frac{R_{c1}}{R_{c3}} \left[1 - \left(\frac{R_{c3}}{R_c} \right)^2 \left(1 + \frac{1}{2} \frac{R_{c3}}{R_{c1}} \right) \right]$$

and since at wavelength λ_0 corresponding to the mid-band frequency f_0, $d = \lambda_0/4$,

$$\cot^2 \frac{2\pi d}{\lambda} = \cot^2 \frac{\pi}{2} \frac{\lambda_0}{\lambda} = \cot^2 \frac{\pi}{2} \frac{f}{f_0}$$

Hence,

$$\cot^2 \frac{\pi}{2} \frac{f}{f_0} = \frac{2R_{c1}}{R_{c3}} \left[1 - \left(\frac{R_{c3}}{R_c} \right)^2 \left(1 + \frac{1}{2} \frac{R_{c3}}{R_{c1}} \right) \right] \tag{3}$$

If the characteristic resistances R_c, R_{c1}, and R_{c3} are known, Eq. 3 provides a means of determining the frequencies f_1 and f_2 below and above the center frequency f_0 at which the main-line sections are perfectly matched. One set of values for the characteristic resistances yields two values for the angle $(\pi/2)(f/f_0)$. From the values of these angles, f_1/f_0 and f_2/f_0 may be calculated.

Waveguide T junctions of the $\mathcal{E}$-plane and $\mathcal{H}$-plane types are illustrated in Fig. 13*A*. Observe that the junctions are formed by inserting a side-arm guide into the broad or narrow face of a straight guide section. Thus

the side-arm guide may be thought of as replacing the cavity in the waveguide-cavity-resonator devices described in Art. 18 of Chap. IX. In these devices a re-entrant cavity, rather than a guide section, is coupled to one of the faces of a rectangular guide.

From a comparison of the junctions of Fig. 13*A* and the equivalent circuits of the waveguide-cavity devices (Fig. 59, p. 694, and Fig. 60,

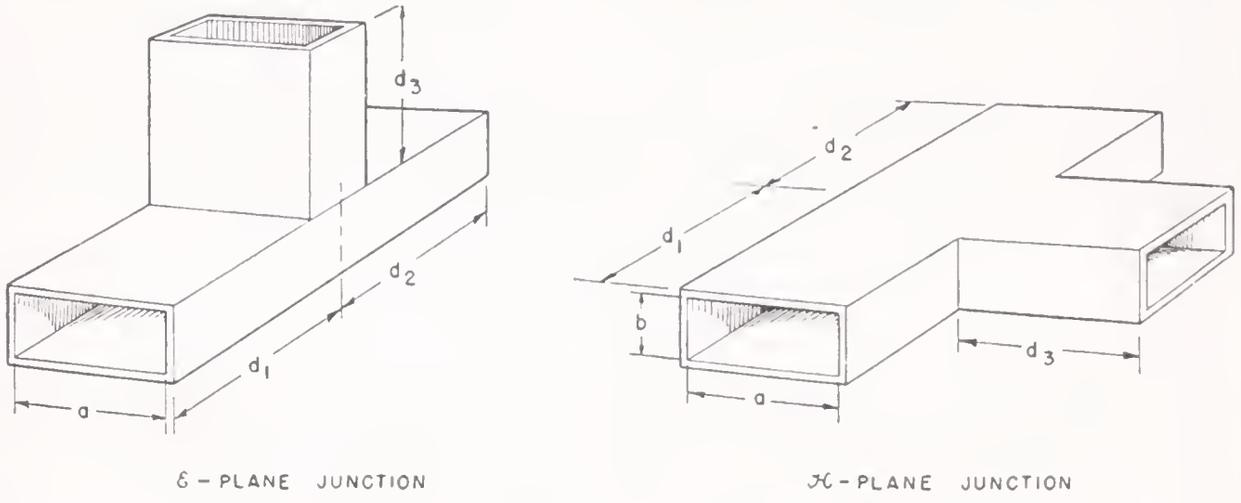


FIG. 13. Waveguide T junctions.

p. 695, of Chap. IX), it may be seen that the ℰ-plane junction provides a series type of connection and the ℋ-plane junction a parallel type of connection. The two forms of connections are shown in Fig. 13*B*. The diagrams of Fig. 13*B* are intended only to indicate the basic types of connections rather than the equivalent circuits of the junctions. If exact equivalent-circuit representations are desired, coupling elements must be inserted at the junction of lines in order to include the effects of discontinuities at the junction.[1]

[1] See, for example, C. G. MONTGOMERY, R. H. DICKE, and E. M. PURCELL, *Principles of Microwave Circuits*, 288–295 (McGraw-Hill Book Company, Inc., New York, 1948).

Of interest in connection with waveguide T junctions are the electric-field patterns which result when a dominant wave is propagated in the side-arm guide of symmetrical $\mathcal{E}$-plane and $\mathcal{H}$-plane junctions, the straight-through sections being terminated in matched loads. As indicated in Figs. 14*A* and *B*, the field in the vicinity of each junction has a rather complicated pattern because of the discontinuities in the guide walls. The composite pattern may, however, be thought of as being the result of both dominant and higher order waves. All wave types present at the junction are coupled to the three component guide sections, but only dominant waves are propagated away from the junction. Thus single traveling waves are present in the straight-through sections, and a

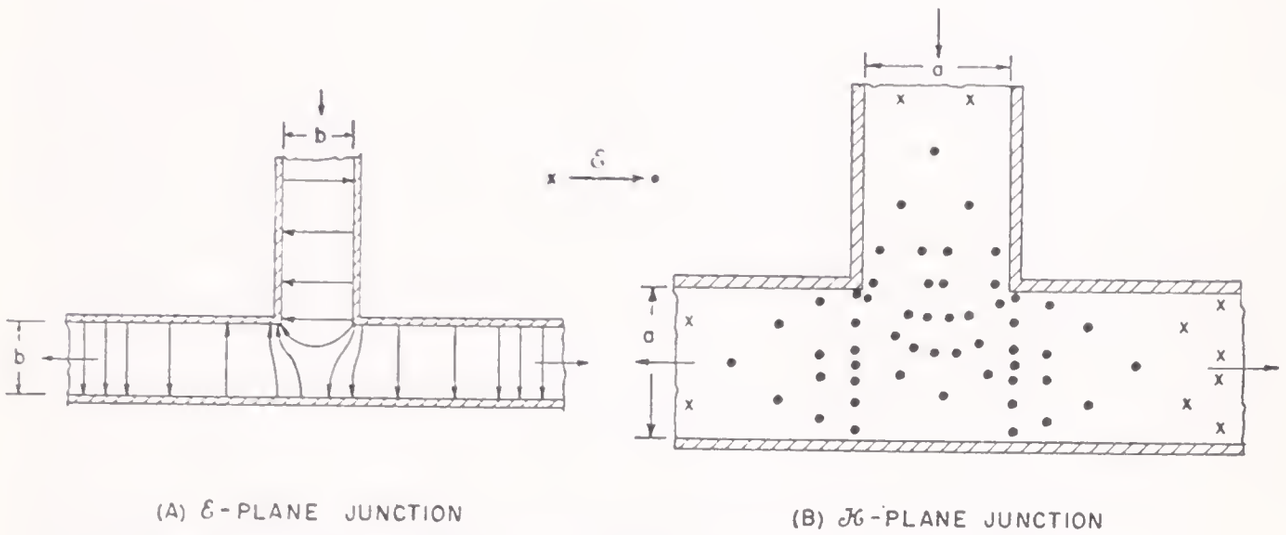


FIG. 14. Dominant-wave electric-field patterns in waveguide T junctions.

partial standing wave appears in the side-arm guide. The partial standing wave results because the straight-through sections do not provide a matched load for the side-arm guide. Observe that the dominant waves in the straight-through sections of the $\mathcal{E}$-plane junction are 180 deg out of phase at the junction of guides whereas in the $\mathcal{H}$-plane junction the waves are in phase at the corresponding point.

6. Hybrid Junctions. A high-frequency device often useful in radar r-f systems is the hybrid junction. A hybrid junction is an essentially lossless device which, at high frequencies, may take the form of a metal enclosure at the junction of *four* transmission lines or waveguides. The hybrid junction guides waves from one to another of the four lines in a particular way. Alternate high-frequency forms of hybrid junctions employ several short lengths of transmission line or waveguide to connect the four main lines. Hybrid junctions built from transformers may also be used to join four low-frequency circuits.

The basic properties of hybrid junctions may be explained by reference to the diagrams of Fig. 15. The first property, illustrated in Fig. 15*A*, is that the junction can be completely matched by resistances; that is, the

values of the four resistors R_1, R_2, R_3, and R_4 may be chosen so that the junction presents to each resistor a resistive input impedance of magnitude equal to that of the terminating resistor. Thus at entry 1 resistor R_1 faces a resistance R_1, at entry 2 resistor R_2 faces a resistance R_2, and so on.

The second property is that, when the junction is completely matched, the entries may be paired in such a way that no power flow exists between entries of a pair. In Fig. 15*B*, the paired entries are 1-2 and 3-4. Thus the upper left-hand diagram of Fig. 15*B* indicates that, with entries 3 and 4 terminated by the resistance values specified in Fig. 15*A*, an input power P at entry 1 divides equally between the loads at entries 3 and 4, and no

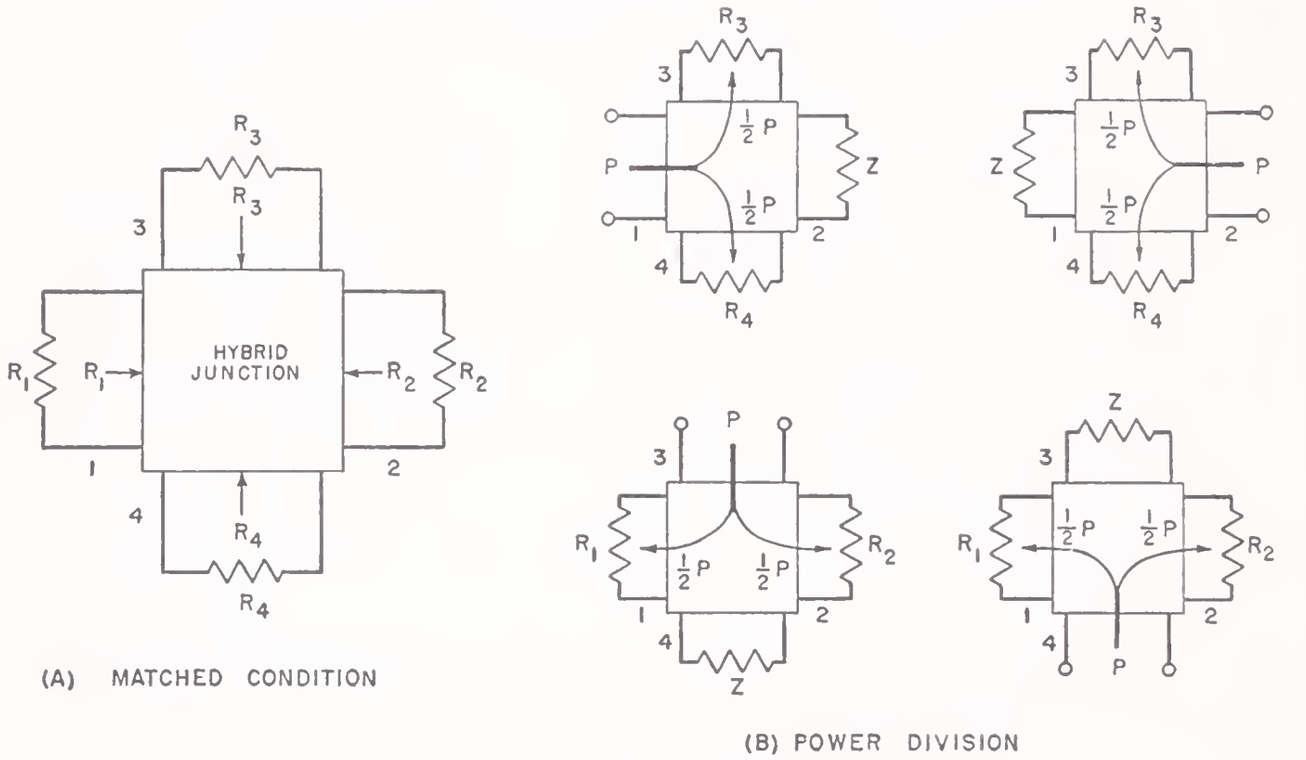


FIG. 15. Properties of hybrid junction.

power reaches the load at entry 2. The power division is independent of the load impedance at entry 2 because no power is delivered to this impedance. The division of power applied at each of the other three entries is indicated in the remaining diagrams of Fig. 15*B*.

The significance of hybrid junction properties may be brought out by the example of Fig. 16, in which power from two generators is coupled through a hybrid junction to two loads. Each entry of the junction is assumed to be connected to a transmission line of characteristic resistance equal to the resistance required at the entry for a perfectly matched junction.

If the lines at entries 3 and 4 are terminated in matched loads, the junction provides matched loads for the lines at entries 1 and 2, and hence no standing waves appear in the system. Furthermore, because entries 3 and 4 are terminated in matched resistances, each generator, if con-

nected in turn in the circuit, delivers equal amounts of power to entries 3 and 4 but no power to the remaining entry. If both generators are connected simultaneously, power is, in general, delivered to entries 3 and 4, but no exchange of power takes place.

With both generators connected, the division of power between entries 3 and 4 depends upon the phase angles of the component voltages appearing at each of the entries. These phase angles depend in turn upon the relative phases of the voltages at entries 1 and 2. The power delivered at each entry must be determined from the resultant voltage and current at the entry rather than by simple addition of the powers delivered when each generator acts alone.

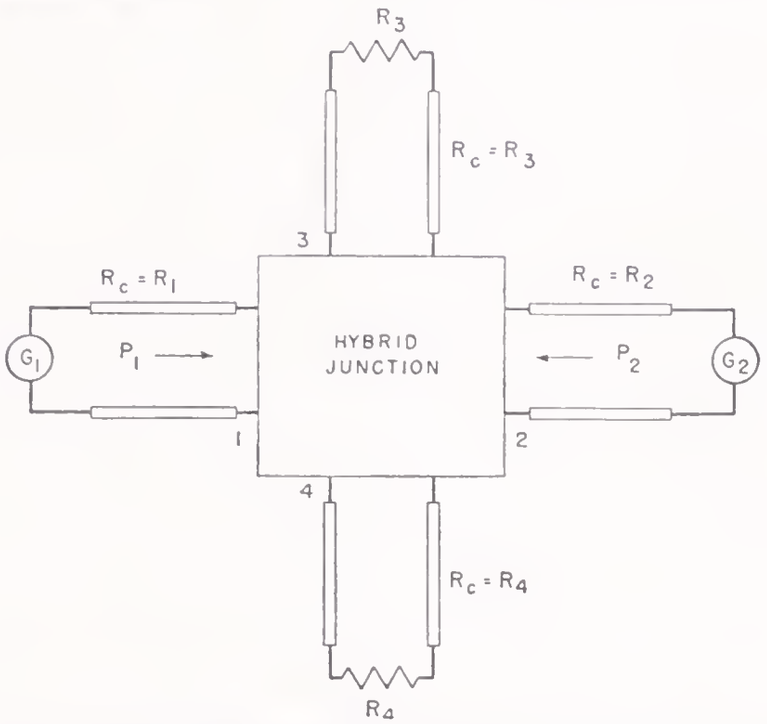


FIG. 16. Use of hybrid junction to connect two generators to two loads.

The power division indicated above is obtained regardless of the nature of the generators. If, however, the generators are linear devices having constant internal impedances equal to the characteristic resistances of the lines to which they are connected, the junction has an additional property of importance. Under this condition of operation, any change in the value of R_4 has no effect upon the power delivered to R_3. A change in the value of R_4 may be thought of as being equivalent to an additional source in the circuit of R_4. Because power from this source divides equally between the matched resistances at entries 1 and 2, the power delivered to R_3 is unaffected. Each load may, therefore, be considered to receive its power from a generator of fixed power output and internal resistance, regardless of the impedance value of the other load. The hybrid junction thus provides power relations very different from those obtained in a parallel connection of two loads to a generator. In the parallel connection, reducing the impedance of one load would in

general reduce the power delivered to the other load because of the generator internal resistance.

The ideal-transformer circuit of Fig. 17 is a hybrid junction that is of interest not only because it can be very closely approximated with actual transformers at low frequencies but also because it serves as an equivalent for one type of high-frequency hybrid junction. That the circuit of Fig. 17 has the properties described for a hybrid junction and that the four equal-valued resistances R are the matching resistances R_1, R_2, R_3, and R_4 for the junction can be shown by direct circuit calculation.

Consider first that power is applied at entry 3, the other entries being terminated in the resistances R. The voltage across the secondary winding of T_1 divides equally between the resistances at entries 1 and 2, and consequently point A is at the same voltage as point B. The input power therefore divides equally between the resistances at entries 1 and 2, and no power is delivered at entry 4. Furthermore, the junction presents a matched resistance to the source at entry 3, since the load resistance across the secondary winding of T_1 is $2R$ and the primary input resistance is $2R$ times $(1/\sqrt{2})^2$, or R. Observe that, if voltage reference directions are taken in the directions indicated in the diagram, the two output voltages e_1 and e_2 are 180 deg out of phase, e_2 being in phase with e_3, and e_1 being 180 deg out of phase with e_3.

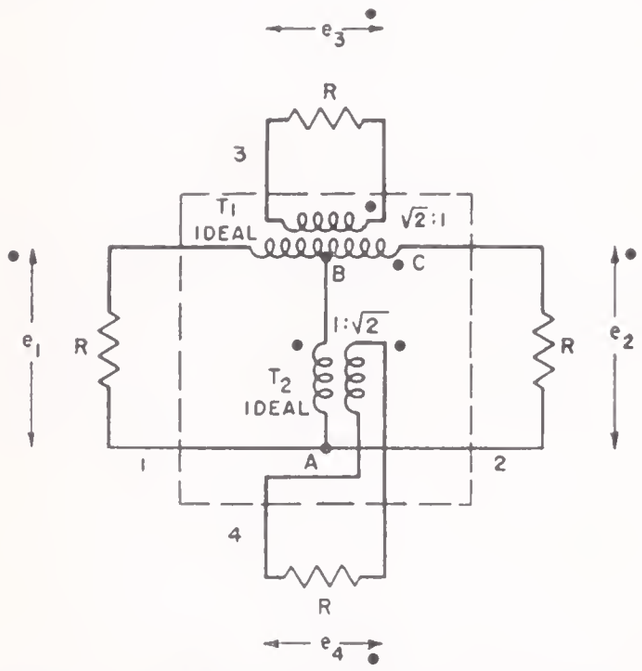


FIG. 17. Ideal-transformer hybrid junction with matching resistors.

If power is applied at entry 4 instead of at entry 3, equal currents flow in opposite directions through the two halves of the tapped winding of T_1. No net magnetomotive force is produced by the winding and consequently no current is required at entry 3. The voltage across both windings of T_1 is therefore zero, and the resistors at entries 1 and 2 are effectively in parallel across the winding of T_2. Thus the input power again divides equally between the loads at entries 1 and 2, and entry 3 is isolated from the source. The input resistance at entry 4 is $(\sqrt{2})^2$ times $R/2$, or R. Note that the two output voltages e_1 and e_2 are now in phase, each being in phase with e_4.

When power is applied to the circuit at entry 1, voltage e_1 divides equally between half the tapped winding of T_1 and the low-voltage winding of T_2, the input resistance for each winding involved being $R/2$. The

voltage between points B and C is equal to the voltage between points B and A; hence no voltage appears at entry 2. The input resistance at entry 1 is R, as before, and the source power divides equally between the loads at entries 3 and 4. The phase of e_4 is the same as that of e_1, whereas e_3 is 180 deg out of phase with e_1 and e_4. Because of the symmetry of the network, similar results are obtained if power is applied at entry 2, the only difference being that the voltages at entries 3 and 4 are both in phase with e_2. The phase relations of voltages in Fig. 17 are not included in the list of general properties of hybrid junctions given at the beginning of the

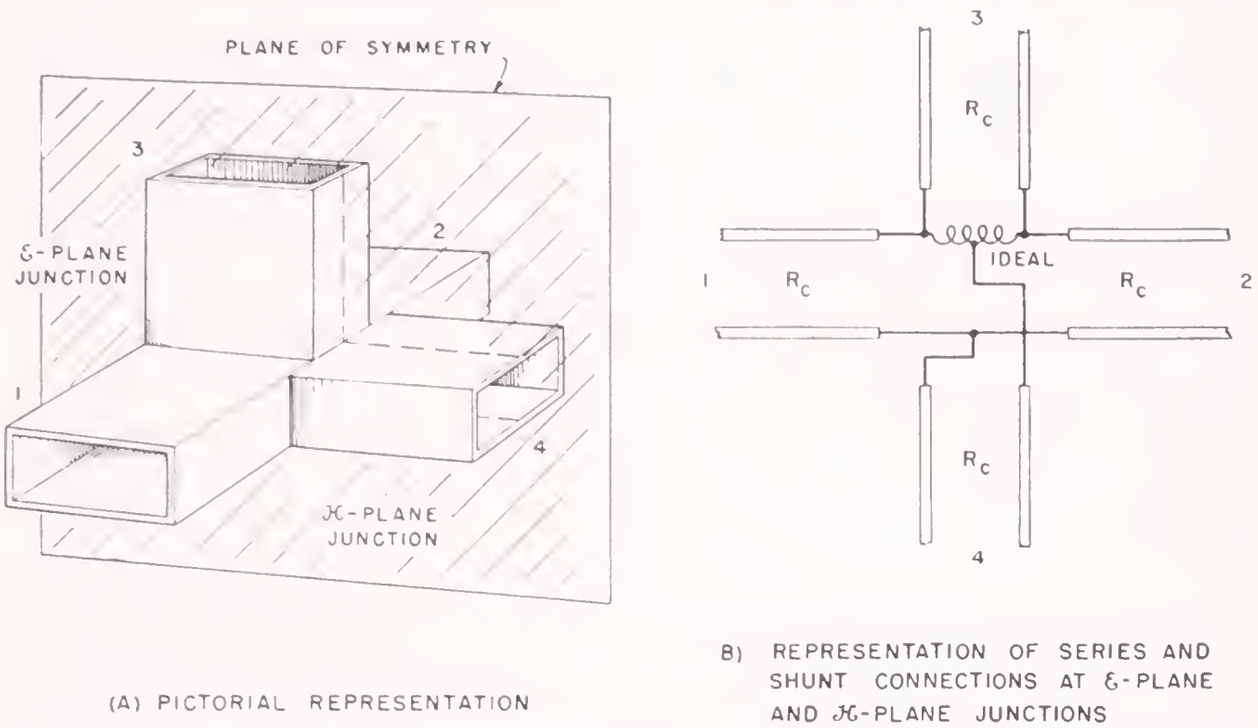


FIG. 18. Waveguide hybrid junction employing combination E-plane and H-plane junctions.

article because different phase relations are obtained in some hybrid junctions.

A high-frequency hybrid junction frequently employed in waveguide systems is a combination of E-plane and H-plane T junctions, as illustrated in Fig. 18*A*. That the device has at least some of the same properties as the circuit of Fig. 17 is evident from a consideration of the symmetry of the fields in the junction. If a $TE_{1,0}$ wave enters guide 3, for example, the electric-field pattern at the junction is somewhat like the pattern in the simple E-plane junction (see Fig. 14*A*, p. 825), and out-of-phase waves emerge from arms 1 and 2. The vertical component of electric field at the entrance to arm 4 is symmetrical about the plane of symmetry of the junction and is upward on one side of the plane and downward on the other. A field of this nature has no tendency to excite a dominant wave in arm 4, since the electric field of such a wave is in the same direction at all points in a cross-sectional plane of the guide. Higher order waves which may be excited are not propagated because guide

dimensions are chosen so that only dominant waves may be transmitted. No power reaches arm 4, therefore, when the source is in arm 3.

Similarly, if power enters arm 4, the electric fields at the junction resemble those of an $\mathcal{H}$-plane T junction (Fig. 14*B*, p. 825), and in-phase waves are propagated in arms 1 and 2. Since the electric fields are in the same direction on either side of the plane of symmetry, there is no tendency to excite a dominant wave in arm 3, and hence the power source in arm 4 delivers no power to arm 3.

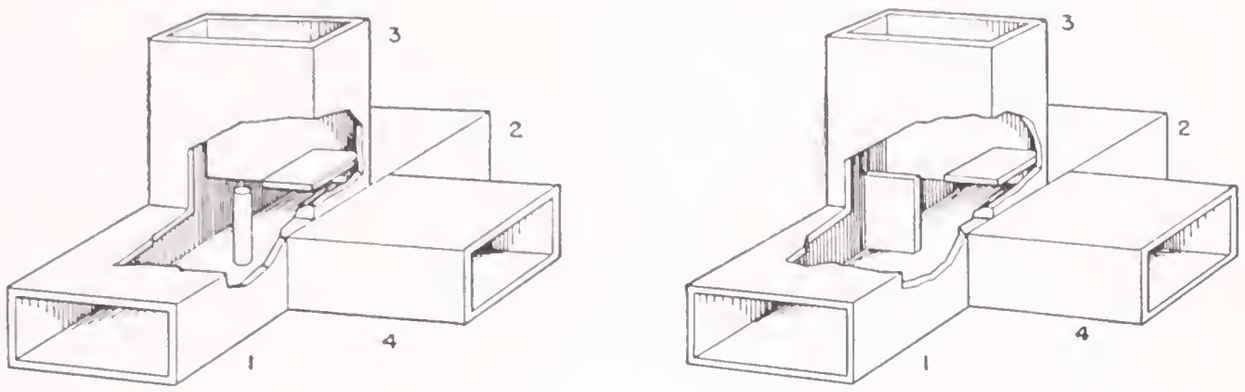
The above properties of the junction of Fig. 18*A* are represented qualitatively in Fig. 18*B*. The waveguides are considered equivalent to transmission lines, and the $\mathcal{E}$-plane and $\mathcal{H}$-plane junctions are represented, in accordance with Fig. 13*B*, p. 824, by series and parallel connections of lines, respectively. The parallel connection is made at the center point of the series connection, a center-tapped ideal autotransformer being introduced to make possible this connection. It should be borne in mind that, because discontinuity effects at the junction are appreciable, Fig. 18*B* is not an accurate equivalent circuit for the junction but only a qualitative representation of the general nature of the connection.

Figure 18*B* suggests that, if the four similar guides are joined as in Fig. 18*A*, and if arms 1 and 2 are terminated in matched loads and power applied through either arm 3 or arm 4, standing waves will be present in the input guide. In fact, if all four lines in Fig. 18*B* have the same characteristic resistance R_c, and if lines 1 and 2 are terminated in matched loads, the terminating impedance for line 3 is $2R_c$ and for line 4, $R_c/2$. Thus a standing-wave ratio of 2 is encountered when either line 3 or line 4 serves as the input line. However, because Fig. 18*B* does not account for discontinuity effects in the guide system, standing-wave ratios in excess of 2 may be expected in the input guide unless matching devices are employed.

In order that the junction of Fig. 18*A* may have all the properties of a hybrid junction when connected to matched guides, standing waves in arms 3 and 4 must be eliminated. Two common forms of matching devices used for this purpose are illustrated in Fig. 19. In order to maintain balance between arms 3 and 4, the matching devices must be inserted in a manner that preserves the symmetry of the junction. The right-hand matching plate in the junctions of Fig. 19 may be considered an asymmetrical iris in guide 3 and is adjusted experimentally in position and width to eliminate standing waves in guide 3. The post (Fig. 19*A*), or the left-hand matching plate (Fig. 19*B*), is then adjusted to produce essentially unity standing-wave ratio in guide 4. Many different forms of matching devices can provide the completely matched condition at a single frequency, but considerable ingenuity is required to devise matching devices that will maintain the match over a broad band of frequencies.

The devices of Fig. 19 can be designed to provide standing-wave ratios less than about 1.2 over bandwidths of the order of 1 per cent in arm 4 and over a considerably greater band in arm 3.

Because the matching devices provide a completely matched junction, they may be considered to add ideal transformers of turns ratios $1:\sqrt{2}$ to the circuit of Fig. 18*B*. Addition of such transformers to Fig. 18*B* converts the junction portion of that circuit into the circuit of Fig. 17. Thus Fig. 17 is an equivalent circuit for the waveguide junction, the four resistances R representing the characteristic resistance of the guides entering the junction. The completely matched waveguide junction has all the properties previously determined for the circuit of Fig. 17. In particular, when each guide is terminated in a matched impedance, not



(A) POST AND PLATE (B) TWO PLATES

FIG. 19. Two forms of matching devices for waveguide hybrid junctions.

only is power entering arm 3 excluded from arm 4, and vice versa, but also power entering arm 1 is excluded from arm 2, and vice versa. The matching devices are necessary to make the waveguide junction of Fig. 18*A* truly a hybrid junction; the name *magic T* is often applied to the matched junction.

A form of hybrid junction that is useful in coaxial-line systems is illustrated in Fig. 20*A* and employs four quarter-wave line sections to connect the four main lines entering the junction. As indicated in the figure, two of the quarter-wave lines have the same characteristic resistance R_c as the main lines and the other two have a lower value $R_c/\sqrt{2}$. The arms are numbered so that entries 1 and 2 form one pair and entries 3 and 4 the other pair, as in Figs. 17 and 18. Observe, however, that the arrangement of entries around the junction is different from that employed in the junctions previously discussed.

An equivalent parallel-wire circuit (Fig. 20*B*), is drawn as an aid to the explanation of the operation of the junction. Assume that power is applied to entry 1 and that entry 2 is temporarily short-circuited, the remaining entries being terminated in R_c, the characteristic resistance of the lines entering the junction. Because the quarter-wave lines *A* and *B*

are each terminated in a short circuit, each line presents an open-circuit at its other termination. The source at entry 1 is therefore effectively connected to the load resistances R_c at entries 3 and 4 through the quarter-wave lines D and C. Since line C is terminated in a matched resistance R_c, the input resistance to line C is also R_c. Line D is therefore terminated by a resistance $R_c/2$, and its input resistance is $(R_c/\sqrt{2})^2/(R_c/2)$, or R_c. Thus standing waves do not appear in the input line connected to entry 1, and the input power divides equally between entries 3 and 4. Because of this power division, voltage E_3 has an amplitude $1/\sqrt{2}$ times that of E_1, and if the positive voltage is in the direction indicated in the diagram, the phase of E_3 is opposite to that of E_1. The currents I_A and

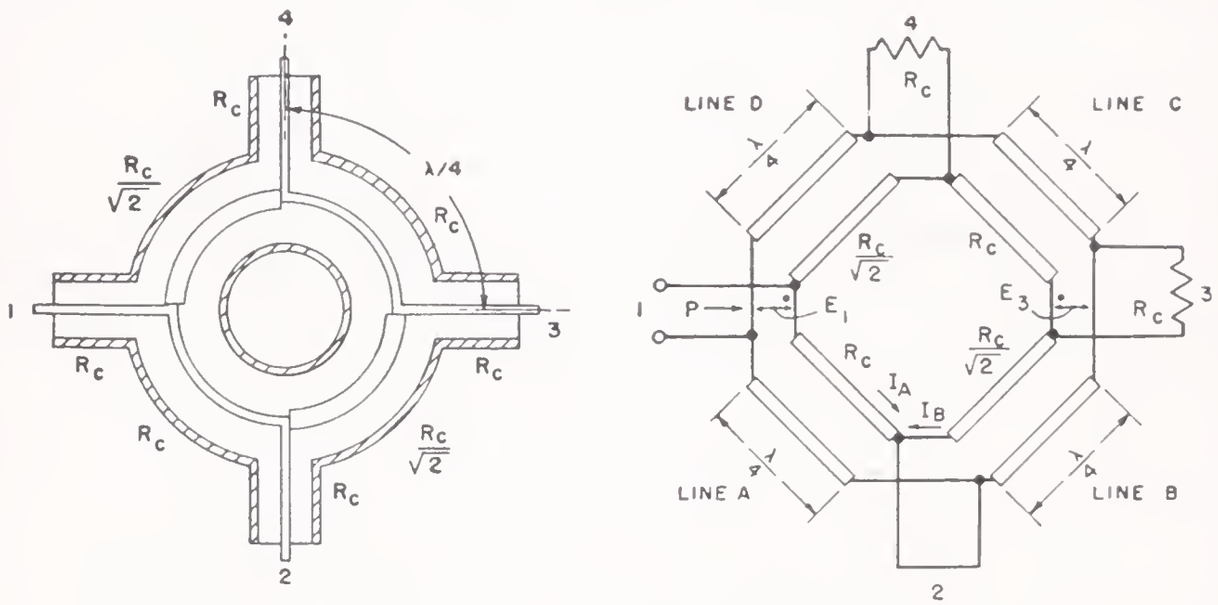


FIG. 20. Coaxial-line hybrid junction of one-wavelength-ring type.

I_B entering the short circuit are equal in magnitude and opposite in phase. (Current I_B has the same amplitude as I_A because the factor $1/\sqrt{2}$ is applied to both the input voltage and characteristic resistance of line B.) The current in the short circuit is therefore zero, and no change in the operation of the circuit results if the short circuit is removed from entry 2.

Because of the symmetry of the ring, the circuit operates in the above manner no matter which entry is connected to the source of power. With matched terminations no power is transferred between entries 1 and 2 or between entries 3 and 4. The device thus has all the properties of a hybrid junction.

A ring type of hybrid junction may be constructed in a variety of ways. For example, parallel-wire lines may be employed in a manner similar to the equivalent circuit of Fig. 20*B*. Alternatively, in the parallel-wire system, series rather than parallel connections may be employed between the ring and the main lines. Use of series connections

requires that lines B and D have a characteristic resistance of $\sqrt{2}\,R_c$. The ring type of hybrid junction may also be employed in waveguide systems, and either $\mathcal{H}$-plane or $\mathcal{E}$-plane T junctions may be used to connect the four main guides to the ring. If connection is made with $\mathcal{H}$-plane junctions, a system equivalent to a transmission-line type of ring with parallel connections at the entries results. Use of $\mathcal{E}$-plane T junctions yields a system corresponding to a transmission-line type of ring with series connections at the entries. In waveguide systems, discontinuity effects at the junctions must be taken into account in determining the proper dimensions of the guides used in the ring.

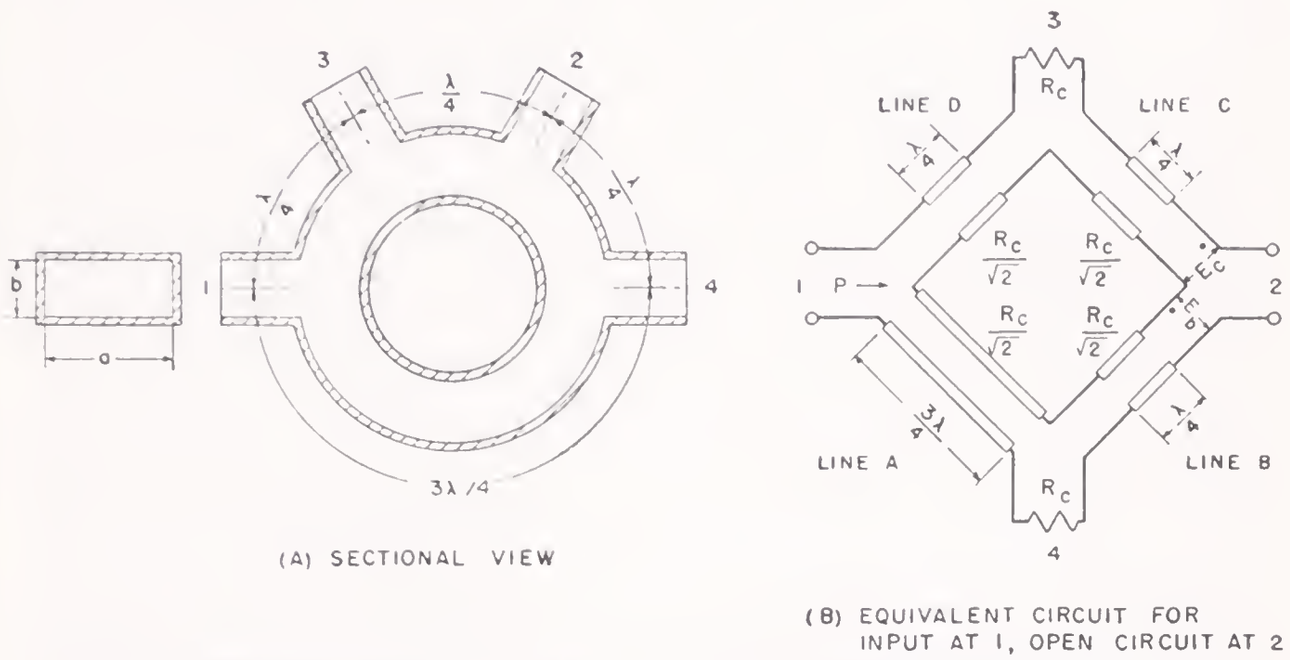


FIG. 21. Waveguide hybrid junction of one-and-one-half-wavelength-ring type.

An alternate form of ring type of hybrid junction differs from the junction of Fig. 20 in that all four sections of the ring have the same characteristic impedance, but not all sections have the same length. In such rings the paired entries occupy alternate positions around the ring rather than adjacent positions, as in Fig. 20. The junction may be used in either transmission-line or waveguide systems, and either series or parallel connections to the ring may be employed.

The alternate form of ring type of waveguide hybrid junction is illustrated in Fig. 21A with $\mathcal{E}$-plane T junctions between ring and main guide. The ring comprises three $\frac{1}{4}$-λ sections and one $\frac{3}{4}$-λ section, λ being the guide wavelength. The b dimension of the guide sections in the ring is made less than the corresponding b dimension of the main guides by an amount that makes the impedance of dominant waves in the ring $1/\sqrt{2}$ times the impedance of dominant waves in the main guide sections.

A transmission-line equivalent of the ring is drawn in Fig. 21B for the conditions that power is applied at entry 1, an open-circuit exists at the junction of guide 2 with the ring, and the transmission lines at entries 3

and 4 are terminated in matched loads. The upper and lower transmission paths between entries 1 and 2 are alike, except for an extra half wavelength of line in the lower path. Voltages E_c and E_b are therefore equal in amplitude and, for the reference directions shown in the figure, are 180 deg out of phase. Thus the output voltage at entry 2 is zero, no power is delivered at entry 2, and with entry 1 as the source, operation is the same regardless of the way entry 2 is terminated. With an open circuit at entry 2, lines B and C have zero input impedance, and hence lines A and D each have an input impedance equal to $(R_c/\sqrt{2})^2/R_c$, or $R_c/2$. Similar reasoning may be applied to show that operation of the circuit is the same when any of the remaining entries acts as the source of power. The device therefore has the properties of a hybrid junction.

7. Directional Couplers. The hybrid junction discussed in Art. 6 is a special form of a more general type of four-entry device known as the *directional coupler*. Directional couplers have properties similar to those of a hybrid junction—each entry of the coupler may be terminated in a matched resistance, and when the coupler is completely matched, the entries may be paired in a way that no power flow takes place between entries of a pair. The difference between a hybrid junction and a directional coupler lies in the amount of coupling that exists between entries that do not constitute a pair. In a hybrid junction, the source power divides equally between the two entries receiving power, but in a directional coupler, there need not be an equal division of the source power. In general, one entry receives some fraction of the total power involved, and the other entry the remaining fraction. Should the source power divide equally, the directional coupler becomes identical with the hybrid junction.

Figure 22 indicates the division of power among the entries of a directional coupler for various locations of the power source. The transmission lines connected to the coupler are assumed to be terminated in matched loads and to face an impedance equal to their characteristic resistance. In Fig. 22*A*, the heavy dashed line indicates that most of the power applied at the input terminals of line 1 is coupled to entry 4 and the light dashed line signifies that only a small amount of power appears at entry 3. The remaining entry receives no power, and hence lines 1 and 2 are completely decoupled. If the source is transferred to line 4, as in Fig. 22*B*, entries 1 and 4 are closely coupled as before, entries 2 and 4 are loosely coupled, and entries 3 and 4 are completely decoupled. The division of power resulting when the source is connected to lines 2 and 3 is indicated in the remaining diagrams of Fig. 22.

Devices that approximate the directional-coupler characteristics just described are frequently employed in r-f transmission lines of radar sets to permit evaluation of the r-f system performance. For example, if the

source connected to line 1 in Fig. 22*A* is the radar transmitter and the load terminating line 4 is the antenna, the power delivered at entry 3 is directly proportional to the transmitter power and hence is a measure of the output power of the radar set. If the small amount of power diverted toward entry 3 is used to excite a high-Q cavity such as an echo box (see Art. 18), the cavity may be tuned to resonance and an indication of the transmitter frequency thus obtained. Details of these applications are discussed in Art. 18.

Because the couplers used in the above measurements are intended to operate over a band of frequencies, a small amount of coupling usually exists between entries of a pair. This coupling results because the

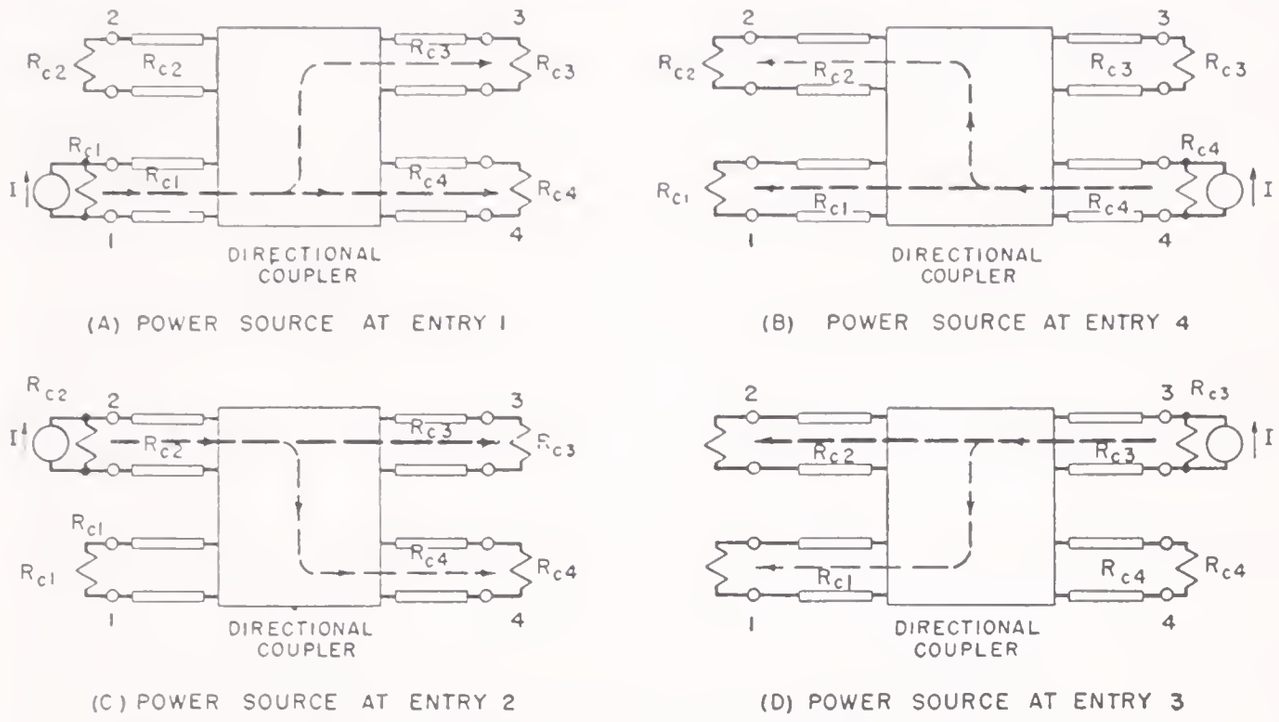


FIG. 22. Power division in a directional coupler.

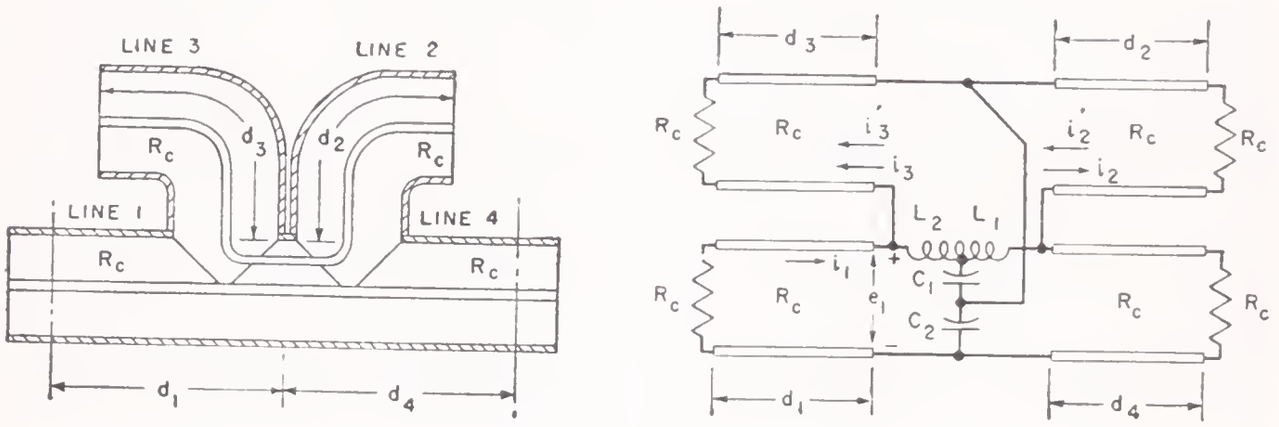
couplers do not provide a perfect match of impedances over the operating range of frequencies. Through careful design, however, it is possible to maintain the amount of this coupling extremely low and thus to approximate closely the ideal coupler.

One type of directional coupler for use with coaxial lines is illustrated in Fig. 23*A*. Branch lines 2 and 3 are sections of coaxial lines, the inner conductors of which form a closed loop inside the mainline sections 1 and 4. All lines are assumed to be terminated in matched loads.

The coupler depends for its operation upon the fact that, when a transverse-electromagnetic wave is propagated in the main-line sections, both the electric- and magnetic-field components of the wave give rise to currents in a circuit comprising the loop and branch lines 2 and 3. In one branch line the current components are in phase and hence permit power to be delivered to the load termination. In the other branch line, the current components are 180 deg apart in phase and thus tend to

cancel. If the branch-line current components are of equal magnitude, no power will be transmitted in one of the branch lines.

The reason that the components of the branch-line currents are in phase in one line and 180 deg out of phase in the other may be explained from Fig. 23*B*, which is an equivalent circuit of the coupler. In terms of the electric field in the main line, the coupling loop may be thought of as constituting a capacitance voltage divider between inner and outer main-line conductors, the voltage between outer conductor and loop being applied to the branch lines *in parallel*. This voltage divider is represented in the equivalent circuit by capacitors C_1 and C_2 in series across the main line. The series inductances L_1 and L_2 in Fig. 23*B* represent magnetic coupling between main line and coupler. In Fig. 23*A* the current



(A) PICTORIAL DIAGRAM (B) EQUIVALENT CIRCUIT

FIG. 23. Coaxial-line directional coupler.

in the main line is associated with a magnetic flux that induces a voltage in series with branch lines 2 and 3. Similarly, in Fig. 23*B*, current in the main line gives rise to a voltage across L_1 and L_2 which is in series with lines 2 and 3. (See Art. 14, Chap. IX).

Assume in Fig. 23*B* that power is applied to line 1 and that all lines are terminated in matched loads. At the instant the main-line voltage e_1 at the coupling loop has the polarity indicated on the diagram, the voltage across C_1 gives rise to branch-line currents i_2 and i_3 in the directions shown. At the same instant the main-line current i_1 at the coupling loop flows to the right in the upper conductor and produces a series voltage which gives rise to branch-line currents i_2' and i_3' in the directions shown. If the current components in each branch line are of equal magnitude, the total current in line 2 is zero and no power is propagated toward entry 2. In line 3, however, the current components add, and power is delivered to the terminating impedance at entry 3.

Actually, zero current in line 2 is not achieved because the field intensities are never exactly uniform along the length of the loop, and as a result, the components of line 2 current are not exactly 180 deg apart in

phase. If the axial length of the loop is made small compared with the operating wavelength, however, a very good approximation to zero current can be obtained.

The magnitudes of the current components in the branch lines may be adjusted by changing the size of the loop and its location between main-line conductors. Increasing the axial length of the loop, while maintaining its position between conductors fixed, increases the amount of magnetic flux linking the loop and the magnitudes of branch-line currents due to the magnetic field are therefore increased. Maintaining the area

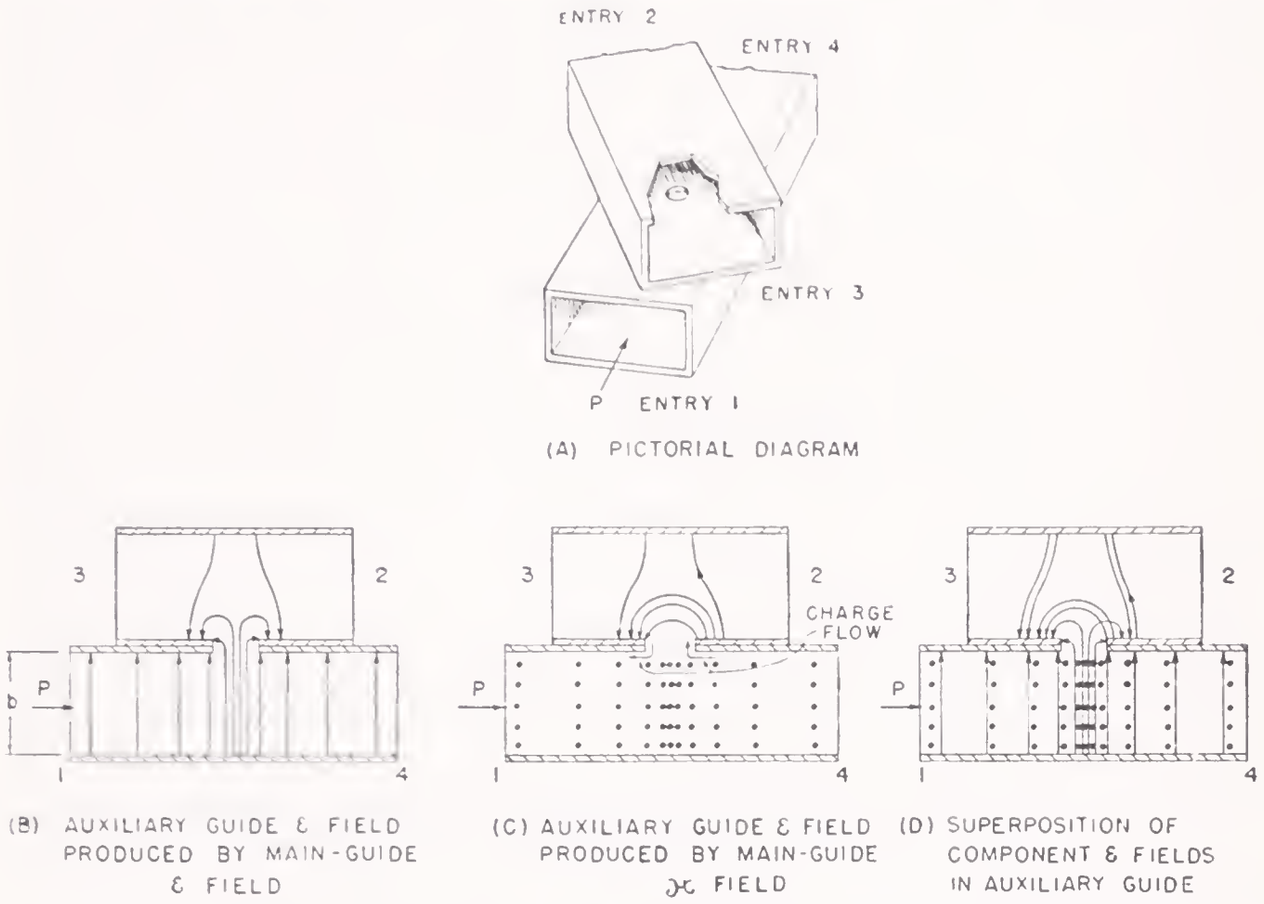


FIG. 24. Single-aperture waveguide directional coupler.

enclosed by the loop constant, while extending the loop farther into the main line, changes the capacitance-voltage-divider ratio so as to increase the output voltage and hence increases the current components due to the main-line electric field.

A waveguide counterpart of the coaxial-line coupler of Fig. 23 may be formed by placing the broad surfaces of two rectangular-guide sections in contact with each other and coupling the two sections together by means of a small aperture in the contacting surfaces. Such a coupler is represented pictorially in Fig. 24*A*.

The directional property of a single-aperture waveguide coupler results from the fact that, when a dominant wave is propagated in the main guide, both the electric field and the transverse component of the magnetic field of the wave produce electric fields in the auxiliary guide. The

directions of the component electric fields in the auxiliary guide are such that the fields tend to cancel in one direction from the aperture and reinforce in the opposite direction. If the magnitudes of the component fields are made equal, a wave will be propagated in one direction in the auxiliary guide but not in the other.

Details of the field patterns in the coupler when a dominant wave is propagated to the right in the main guide are given in Figs. 24*B*, *C*, and *D*. Figure 24*B* shows the electric field produced in the auxiliary guide by the portion of the main electric field that fringes through the aperture, and Fig. 24*C* indicates the electric field set up in the auxiliary guide by the main magnetic field. The latter component of electric field results because the coupling aperture interrupts a portion of the longitudinal current associated with the transverse magnetic field in the main guide. When the two component electric fields in the auxiliary guide are superposed, as in Fig. 24*D*, it may be seen that the fields to the left of the aperture reinforce one another and give rise to a dominant wave which travels to the left in the upper guide. The fields to the right of the aperture tend to cancel, however, and if an exact cancellation is achieved, no power is transmitted to the right. Power transmission to the right may be minimized through proper shaping of the coupling aperture or by skewing the axis of the auxiliary guide with respect to the main-guide axis (see Fig. 24*A*).

When the directional coupler of Fig. 24 is used for making measurements on the r-f system of a radar set, one end of the auxiliary guide is terminated in an absorbing material that provides a matched load for the guide section. Details of this type of termination are discussed in the next article. The opposite end of the auxiliary guide contains a pickup probe and output device. These elements also provide, as nearly as possible, a matched load for the auxiliary-guide section.

A waveguide directional coupler employing two coupling apertures is illustrated in Fig. 25*A*. The guide sections are joined along their narrow walls and coupled together by means of two similarly shaped apertures spaced one-quarter guide wavelength apart.

Because the holes are located at points of zero electric field in the main-wave pattern, coupling between guides is due almost entirely to the magnetic-field component of the main wave. Thus each coupling aperture gives rise to two waves—one in each direction in the auxiliary guide. The directional property of the coupler results because of cancellation of the component waves traveling in one direction and reinforcement of the component waves propagating in the opposite direction.

Details of the operation of the coupler of Fig. 25*A* may be explained in terms of its two-wire equivalent circuit drawn in Fig. 25*B* in accordance with the principles explained in Art. 18 of Chap. IX. In the equivalent

circuit, inductances L_a and L_b account for the coupling provided by each aperture, and resistors R_1, R_2, R_3, and R_4 provide matched loads at the various entries. When a wave is propagated to the right in the main line, a small portion of the power in the wave is diverted at point A toward the auxiliary line, and as a result, two equal-amplitude waves traveling in opposite directions are set up in the auxiliary line. Similarly, a small portion of the main-line power is diverted at point B toward the auxiliary line, and two more waves traveling in opposite directions on the auxiliary line are set up. Because the component waves traveling to the right of point B on the auxiliary line traverse paths of equal length, they are in

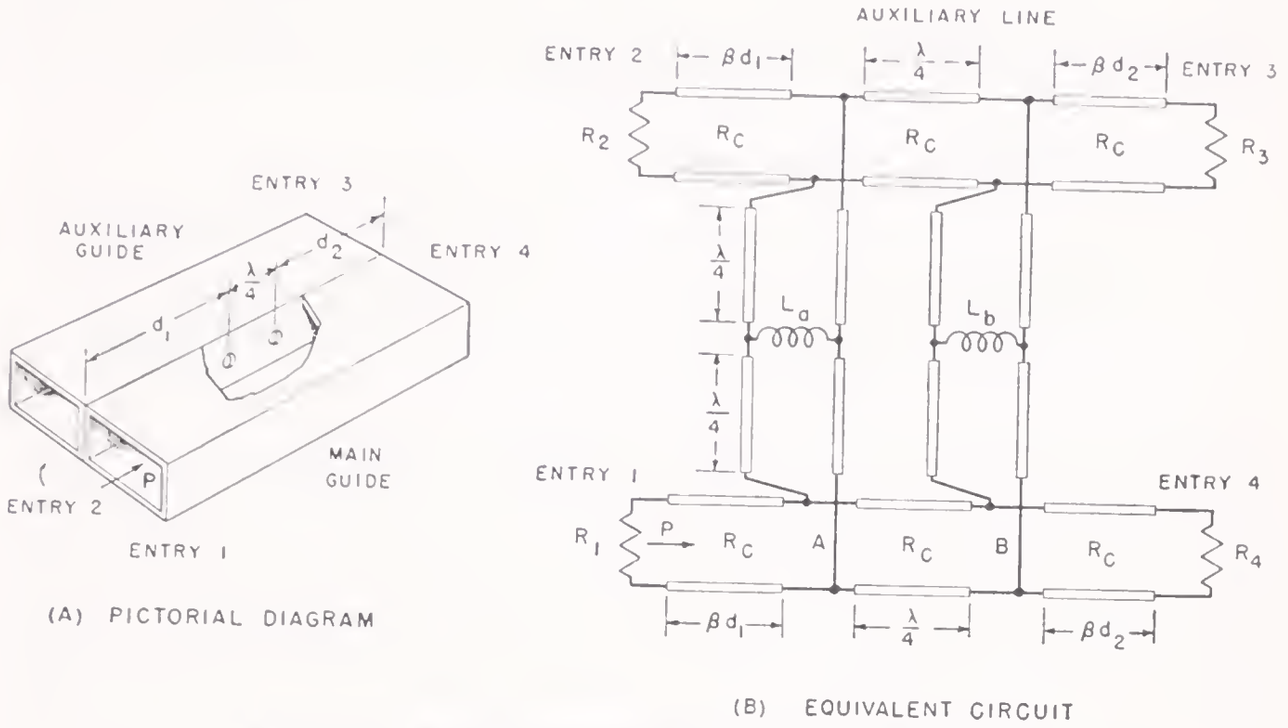


FIG. 25. Two-aperture waveguide directional coupler.

phase and hence reinforce each other. Power is therefore delivered to R_3. The component waves traveling to the left of point A on the auxiliary line are 180 deg apart in phase, however, since they traverse paths that differ in length by one-half wavelength. These waves therefore cancel, and no power is transmitted to R_2. If the coupling inductances L_a and L_b are small, the input impedance of the branch lines in parallel with the main line is extremely high; consequently, very little of the main-line power is diverted to the auxiliary guide, and very little mismatch is introduced into the main line by the coupling apertures.

Two-aperture coaxial-line directional couplers are similar to two-aperture waveguide couplers, except that the aperture spacing is made one-quarter wavelength as measured in space. Observe that in two-aperture couplers power flow in the auxiliary line or guide section is in the same direction as in the main section, whereas in single-aperture couplers, power flows in the main and auxiliary sections in opposite directions.

8. Attenuators. Frequently, in r-f transmission systems and associated test equipment, attenuators are inserted at points of the r-f line or waveguide in order to reduce the amount of power transferred at the attenuator location. Local oscillators, for example, are sometimes connected to the main transmission line through an attenuator (see Art. 16), and r-f power-measuring devices usually include an attenuator in their input circuits to limit the power applied to the measuring circuit. In still another application, an attenuator is required to absorb without reflection all the power incident upon it and thus to provide a matched-load termination for the line or guide to which it is connected. At low-power levels matched-load terminations are required at one end of the auxiliary line or guide of a directional coupler. At high-power levels they are used to replace the antenna of a radar system during periods of testing.

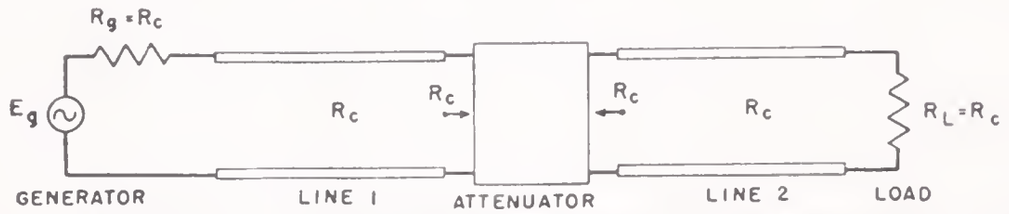


FIG. 26. Attenuator inserted between two matched transmission-line sections.

Attenuators may control the amount of power transferred between points of a transmission system either by absorbing some of the power in the system or by reflecting some of it. Attenuators that operate on the power-absorption principle use, as the dissipative element, materials such as finely divided iron or carbon, a high-resistance alloy such as nichrome, various types of high-loss dielectric materials, or extremely thin layers of low-resistance metals. Attenuators that operate on the power-reflection principle employ sections of waveguide operated at frequencies below cutoff. Both types of attenuators are discussed in this article.

The amount of attenuation that results when an attenuator is inserted between two sections of transmission line is usually defined for the impedance conditions indicated in Fig. 26. The line sections have a characteristic resistance R_c, and the generator and load impedance are assumed to be equal to R_c. For these impedance conditions, the attenuator is made to have an input and output impedance equal to R_c; hence, the generator and load each face an impedance R_c. In absence of the attenuator, the generator delivers maximum power to the load, the amount of power being $P_1 = E_g^2/4R_c$. The reduction in power to the load, or the attenuation A in decibels, that results when the attenuator is inserted in the line is

$$A = 10 \log \frac{P_1}{P_2}$$

where P_2 is the power delivered to the load with the attenuator in the line.

Observe that, because the generator and load face matched impedances, the attenuation A is a property of the attenuator alone and is independent of the lengths of the line sections. If, on the other hand, the input and output impedances of the attenuator are such that the generator and load do not face matched impedances, the attenuation depends not only upon the attenuator characteristics but also upon line lengths and is thus more sensitive to slight changes in frequency. Microwave attenuators are usually calibrated under the matched-impedance conditions illustrated in Fig. 26. In order for the calibration to be significant, therefore, it is necessary that the attenuator be employed in lines or guides that provide similar conditions of impedance match.

Power-absorbing Attenuators. The simplest form of coaxial-line attenuator of the power-absorbing type is a coaxial cable having a high-resist-

(A) STUB-SUPPORTED INNER CONDUCTOR (B) BEAD-SUPPORTED INNER CONDUCTOR

FIG. 27. Coaxial-line attenuators having metalized-glass inner conductors.

ance inner conductor. One type of such cable, designated as Type RG-21/U, makes use of nichrome wire as the inner conductor and has an attenuation of approximately 0.8 db per ft at 10 cm and 1.6 db per ft at 3.3 cm. The cable is therefore of convenient length when a few decibels of attenuation are required but becomes unwieldy when large amounts are needed. High-loss coaxial cable is used primarily in applications where precise values of attenuation and good impedance matching are not extremely important. The attenuation of a fixed length of cable may vary appreciably with changes in temperature and may also be affected if the cable is subjected to bending or twisting movements. Although the characteristic impedance of the RG-21/U cable is nominally 50 ohms and is very nearly resistive, the actual characteristic impedance may differ from the nominal value by as much as ± 5 per cent. Thus if the cable is connected to a 50-ohm line, the resulting mismatch of characteristic impedances may give rise to standing-wave ratios as high as 1.1 on the 50-ohm line.

A more precise type of coaxial-line attenuator than the high-loss coaxial cable is illustrated in Fig. 27. As indicated in the figure, these attenuators make use of a section of air-filled, rigid, coaxial line, the inner conductor of which consists of a thin film of metal on a dielectric support,

usually glass. Either a stub support or beads may be used to hold the inner conductor in place. The inner conductors are constructed as separate units and provided with bullet-shaped end connectors that fit into the beads or stub. This type of fixed attenuator, or *pad* as it is sometimes called, is quite rugged and can be expected to maintain its calibration despite changes in temperature and humidity.

The metalized-glass attenuators of Fig. 27, like high-loss cables, make use of high-resistance inner conductors. The inner-conductor resistance is made high by maintaining the thickness of the metallic film less than the skin depth of the film material. Actual film thicknesses normally range from 10^{-6} to 10^{-5} cm, depending upon the amount of resistance required to produce the desired attenuation. The nature of the film material likewise depends upon the amount of resistance required and also upon the process used to deposit the material on the support.[1] Commonly used film materials are nichrome and mixtures of platinum and gold or palladium.

In order that metalized-glass attenuators may be of convenient length and still provide large amounts of attenuation, the per-unit-length resistance of the attenuator must be appreciably larger than that of a low-loss line. The characteristic impedance of the attenuator may therefore be expected to contain an appreciable reactive component, and as a result, reflections may be expected at the ends of the attenuator unless suitable matching sections are provided. Usually a matching section is located at each end of the attenuator. The section nearer the power source serves to provide a matched load for the line section between source and attenuator, and the section nearer the load ensures that the output impedance of the attenuator matches the line section between attenuator and load. The matching sections are made an integral part of the attenuator by changing the film thickness of the inner conductor (and hence the characteristic impedance of the attenuator) by the proper amount near each end. In units that provide less than about 10 db of attenuation, the matching sections may be omitted without causing serious mismatches of impedances, but for attenuations appreciably greater than 10 db, matching sections are necessary.

Metalized-glass rods may also be used in variable coaxial-line attenuators. In one type a metal tube is arranged to slide over the metalized-glass rod forming the inner conductor of the attenuator. A portion of the metallic film is thus short-circuited by the tube, and variable amounts of attenuation are obtained. Attenuators which provide from 3 to 40 db of attenuation with standing-wave ratios less than 1.25 over a ± 10 per cent

[1] A very thorough discussion of the techniques involved in fabricating metalized-glass attenuators is contained in C. G. MONTGOMERY, *Technique of Microwave Measurements*, 757–762, 774–778 (McGraw-Hill Book Company, Inc., New York, 1947).

frequency band are possible with this type of construction. A similar performance, except for slightly higher standing-wave ratios, may be obtained by allowing the metal tube to slide inside the metalized-glass rod. The glass provides high-capacitance coupling between the metal tube and metallic film and effectively short circuits the film to the tube.

Figure 28*A* illustrates a waveguide attenuator of the resistive type, formed by inserting a metalized-glass plate in the guide parallel to the side wall. The metal film short-circuits a portion of the electric field in the guide and gives rise to a current in the film. The film current results in a portion of the guide power being dissipated. The film is deposited on one side of the plate and covers the entire surface except for small rectangular sections near the ends (see Fig. 28*A*). The metallic end sections thus formed provide the necessary impedance matching between

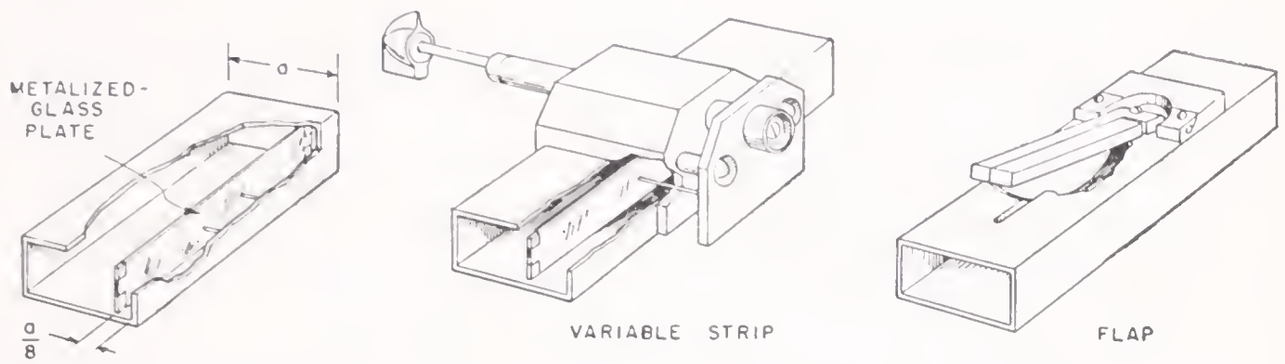


FIG. 28. Waveguide attenuators.

attenuator and guide sections connected to it. The dimensions of the main attenuating film and matching sections are determined by the amount of attenuation required.

The glass plate is supported within the guide by means of two metal rods cemented to the plate and attached to the side walls of the guide. The plate is usually located a distance of approximately $a/8$ from one side wall, where a is the guide width. Locating the plate at this point yields an attenuator relatively insensitive to changes in frequency. An alternative to the metalized-glass plate is a resistance card made by coating a thin sheet of dielectric material with a mixture of graphite and a suitable binder. Resistance-card attenuators are in general more sensitive to changes in temperature and humidity than are the metalized-glass type.

The attenuator of Fig. 28*A* may be made variable by moving the plate in a way that allows the metallic film to short-circuit various amounts of electric field. In one type of variable attenuator (Fig. 28*B*, left), the plate is moved back and forth across the guide by means of a drive mechanism. An alternative type (Fig. 28*B*, right) moves the plate in and out of the guide at a fixed location between side walls.

Matched-load Terminations. Power-absorbing attenuators having high values of attenuation make suitable terminations for a transmission line or waveguide, provided they are able to dissipate the required amount of power without danger of burnout. In some applications, as in directional couplers, the power involved is very small, being of the order of a few tenths of a watt. In other applications, such as when the attenuator is used as a load termination for a radar transmitter, as much as 1 kw of average power must be absorbed.

Low-power coaxial-line loads are constructed, as shown in Fig. 29, by closing off the end of a line section with a metallic plug and inserting a power-absorbing material in the space near the end of the section. The power-absorbing material may be a high-loss dielectric cloth impregnated with a power-dissipating material such as finely divided carbon and a

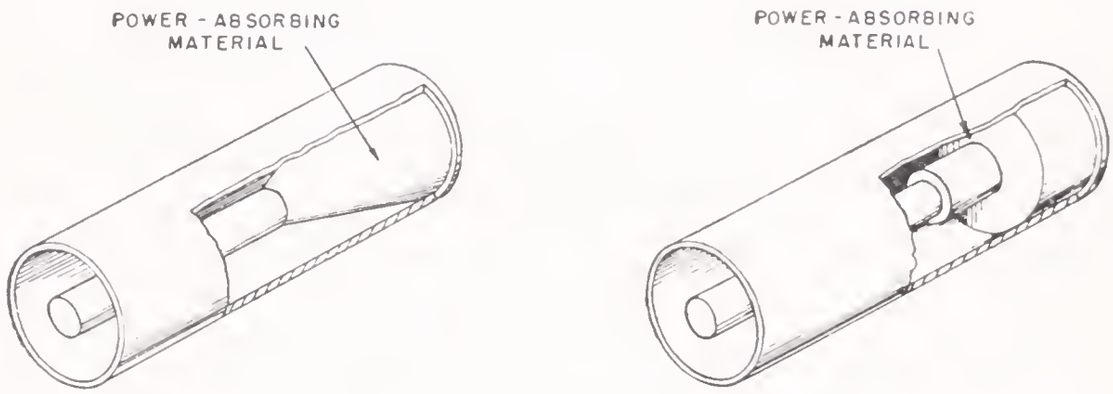


FIG. 29. Low-power matched-load terminations for coaxial lines.

binder or finely divided iron powder and a binder. The binder may be bakelite or a form of polystyrene. A good impedance match is obtained between load and line either by tapering the dissipative material as shown in Fig. 29*A* or providing it with a shoulder as indicated in Fig. 29*B*. Standing-wave ratios less than 1.04 can be attained over a 10 per cent frequency band with terminations of this type.

Low-power loads for rectangular guides employ as the power-absorbing material either powdered iron and a binder or a strip of dielectric material on which is deposited a film of resistive material such as finely divided carbon mixed with a binder. For 3- and 1-cm guides a low-reflection termination may be obtained by tapering a block of powdered-iron material toward the bottom wall and one side wall of the guide, as shown in Fig. 30*A*. Usually the cross-sectional dimensions of the uncut block are made slightly less than those of the guide. An alternative form for the block is illustrated in Fig. 30*B*. With either shape, the standing-wave ratio in the main guide can be maintained less than 1.02 over a frequency band of several per cent.

At 10 cm a resistance card is preferable to an iron block because of the excessive size of block required. The card is located parallel to the side walls of the guide and extends from top to bottom wall. The back edge of the card is placed against the closed end of the guide, and the front edge is tapered from top and bottom walls to the axis of the guide as a means of providing impedance matching. The over-all length of the card is approximately one guide wavelength. A somewhat shorter and

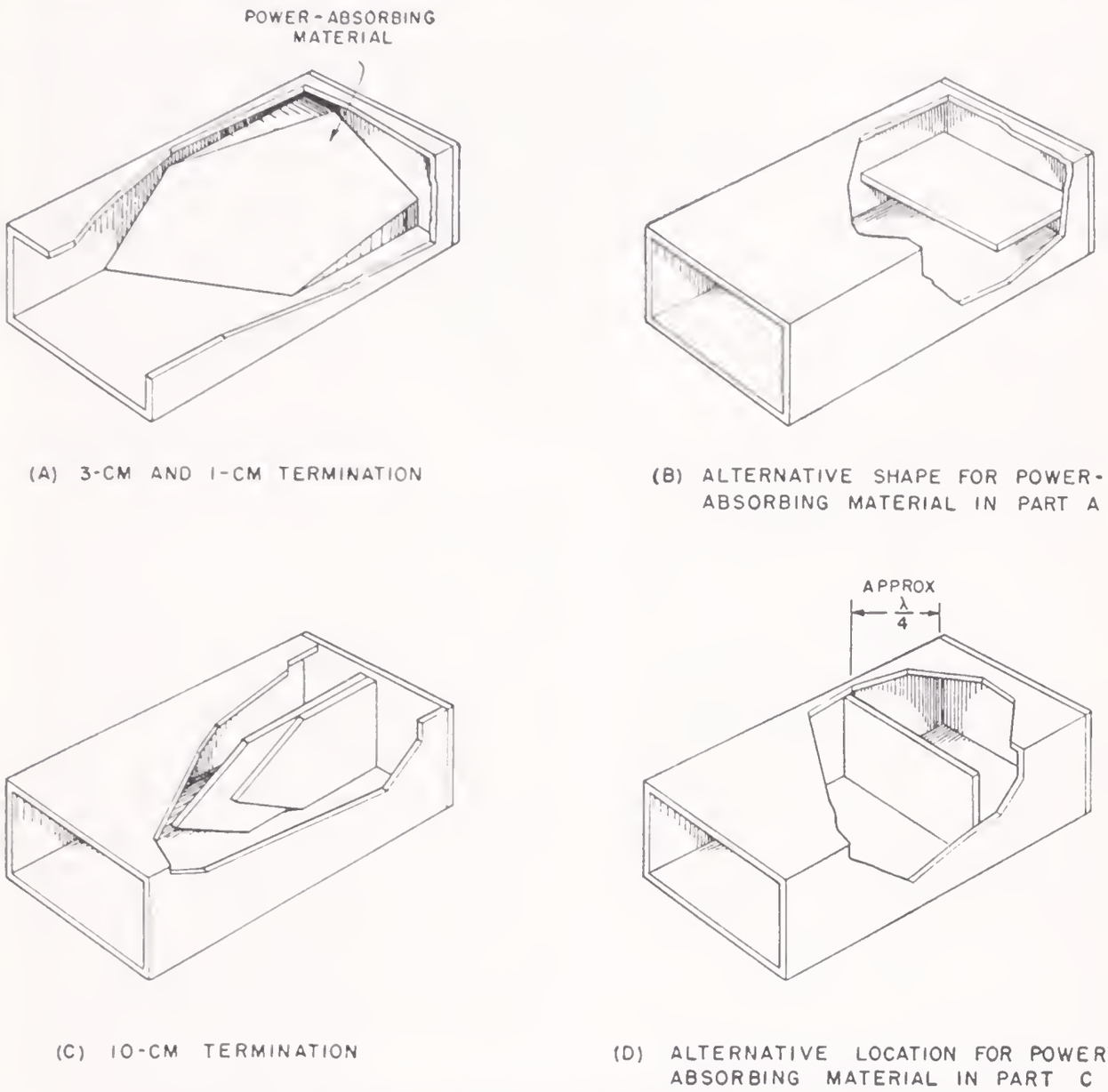

POWER-ABSORBING MATERIAL
(A) 3-CM AND 1-CM TERMINATION
(B) ALTERNATIVE SHAPE FOR POWER-ABSORBING MATERIAL IN PART A
APPROX
λ/4
(C) 10-CM TERMINATION
(D) ALTERNATIVE LOCATION FOR POWER-ABSORBING MATERIAL IN PART C

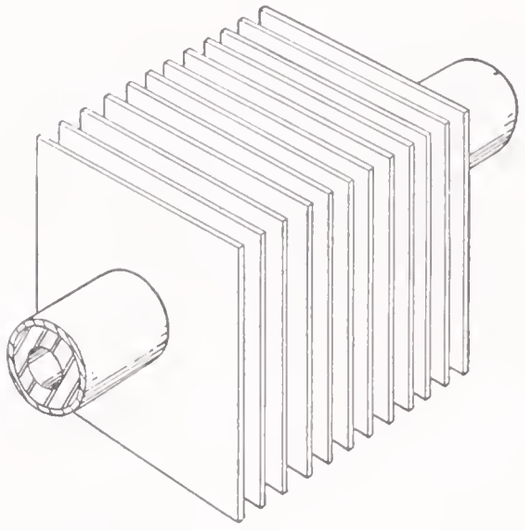
FIG. 30. Low-power matched-load terminations for waveguides.

more rugged termination may be obtained by placing two cards side by side, as in Fig. 30*C*.

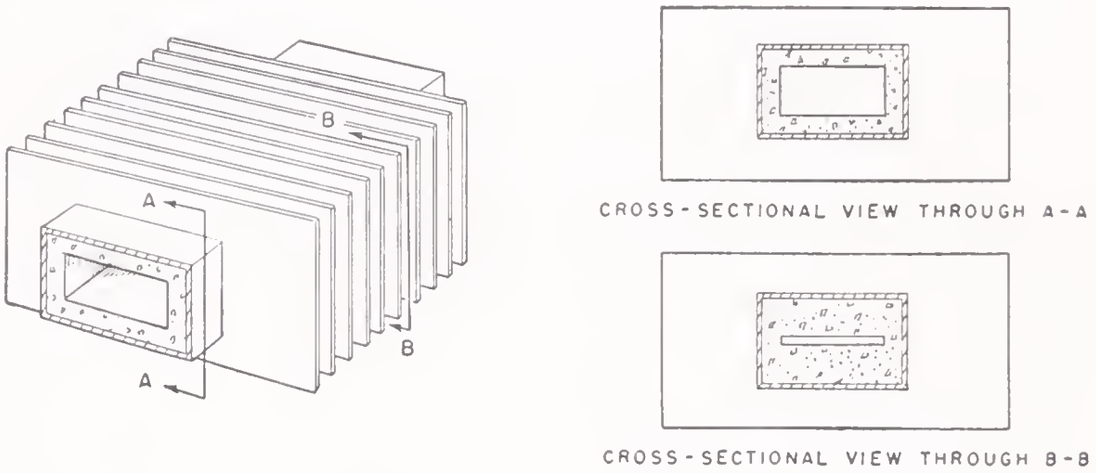
If standing-wave ratios of the order of 1.2 can be tolerated, the resistance card may be placed at right angles to the axis of the guide, as shown in Fig. 30*D*. It is usually desirable to close off the guide approximately a quarter wave beyond the card in order to make the portion of guide to the rear of the card the equivalent of an open circuit in parallel with the card resistance.

Instead of resistance cards, metalized-glass plates similar to those illustrated in Fig. 27 may be employed in low-power load terminations for waveguides. Metalized plates have the advantage of being more rugged than resistance cards and are less affected by changes in temperature and humidity.

High-power matched-load terminations for coaxial lines and waveguides are pictured in Fig. 31. The coaxial-line load (Fig. 31*A*) is



(A) COAXIAL - LINE LOAD



(B) RECTANGULAR - WAVEGUIDE LOAD

FIG. 31. High-power coaxial-line and waveguide loads.

intended for 10-cm operation and makes use of powdered graphite and a cement binder as the power-absorbing material. The mixture is molded and is tapered toward the input end to provide a matched termination. The fins attached to the outside of the assembly serve as extra cooling surfaces. Loads of the type illustrated in Fig. 31*A* are capable of dissipating up to 100 watts of average power and of maintaining in the lines they terminate a standing-wave ratio of less than 1.1 over a ± 10 per cent frequency band.

High-power waveguide loads are commonly constructed by making the walls of the guide from a power-absorbing material similar to that used in the coaxial-line load. The external appearance and constructional details of such a load are given in Fig. 31*B*. In order to ensure uniform dissipation of power throughout the load as well as good impedance matching, the *b* dimension of the load is diminished gently near the input end and then rather sharply as the far end is approached. Only a slight reduction is made in the *a* dimension with distance along the axis of the load. With this type of construction 10-cm waveguide loads capable of dissipating up to 1 kw of average power are possible. Similar 3-cm loads can handle up to 200 watts of average power.

Cutoff Attenuators. A cutoff type of attenuator depends for its operation upon a waveguide operating below cutoff. Because the fields in such a guide attenuate exponentially with distance along the axis, it is possible to vary the amount of power transferred between input and output ends of the guide simply by varying the guide length.

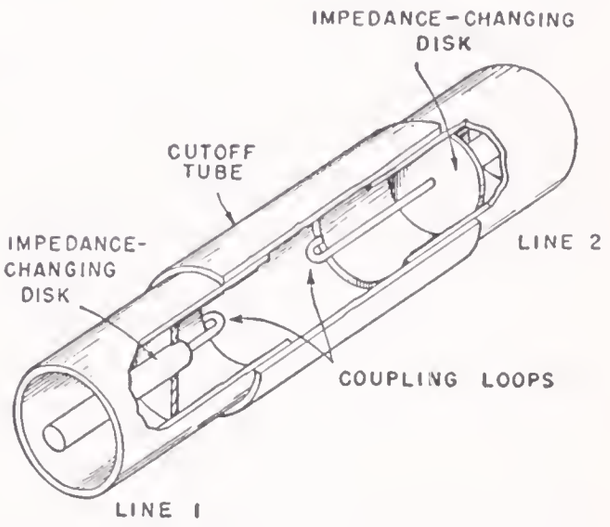


FIG. 32. Basic features of a cutoff type of attenuator.

The essential features of a cutoff attenuator are illustrated in Fig. 32. A section of waveguide in the form of a hollow metal tube is placed over the outer conductors of two rigid coaxial-line sections, the guide diameter being chosen so that operation is below the cutoff frequency. The line sections are terminated within the guide by coupling loops separated a distance which depends upon the amount of attenuation required. (In variable attenuators, one line section is rigidly attached to the guide and the other section is allowed to slide back and forth within the guide.) The circular disk shown in each line section serves as an impedance-matching device.

When a power source is connected to line 1, the current in the line-1 coupling loop gives rise to a magnetic field in the form of closed loops along the axis of the cutoff tube. The magnetic-field pattern is similar to a single half-wave pattern of a $TE_{1,1}$ wave in a circular guide except that the intensity of the field diminishes with distance from the coupling loop terminating line 1. The intensity of the magnetic field linking the line-2 coupling loop and hence the magnitude of the voltage induced in the loop depend upon the separation distance of the two coupling loops.

An electric field is also present within the cutoff tube, and as explained in Art. 12 of Chap. VIII, the electric and magnetic fields are 90 deg apart

in phase. The impedance associated with the fields is therefore in the nature of a reactance.

An advantage of a cutoff type attenuator is that the amount of attenuation resulting from a given separation distance of the coupling loops may be determined by direct calculation. From Eq. 65 of Chap. VIII, α, the below-cutoff attenuation per unit length of guide, is

$$\alpha = \frac{2\pi}{\lambda_0}\sqrt{\frac{\lambda_0}{\lambda_{co}} - 1} \tag{4}$$

where λ_0 is the space wavelength and λ_{co} is the cutoff wavelength of the traveling wave which corresponds to the field pattern in the guide. If the pattern corresponds to that of a $TE_{1,1}$ wave, λ_{co} equals 1.71 times the diameter of the guide. Multiplying Eq. 4 by the separation distance of the loops gives the total attenuation.

In general, calculated values of attenuation are most representative of actual values in the region of high attenuation. In the low-attenuation region, the fields within the cutoff guide are likely to include one or more components that correspond to higher order wave types. Each component field may contribute coupling between input and output loops, in which case the over-all attenuation depends upon the relative amplitudes and phases of the various component fields at the output loop. Because the higher order fields attenuate more rapidly than the $TE_{1,1}$ field, the effect of higher order fields becomes less important as the loop separation distance becomes large and is generally negligible for separation distances that yield attenuations above 20 db.

Impedance-changing devices are always included in cutoff attenuators because of the large amount of mismatch that exists between the line sections and waveguide. At high values of attenuation, the coupling between lines 1 and 2 is extremely small, and the terminating impedance of line 1 is essentially the reactive impedance of the below-cutoff guide. In order to provide a matched load for line 1, therefore, a complex-impedance element must be located near the terminating loop. A similar element is required in the output line to make the output impedance of the attenuator match the characteristic resistance of the output line.

One form of impedance-changing device commonly employed makes use of a circular resistive disk normal to the axis of the line (see Fig. 32). The disk is constructed from a dielectric material and carries a coating of carbon on one surface. Rings of silver paste baked on the disk at the inner and outer edges permit the disk to be soldered in place. At microwave frequencies the disk is equivalent to a complex impedance across the line, the resistive component of the impedance being due mainly to the carbon material, and the reactive component to the combined effects of the capacitance resulting from the presence of the dielectric material and

the inductance of the carbon surface. By properly locating the disk along the line it is possible to cancel the reactive component of the disk impedance with the reactive component of the input impedance of the attenuator and to present a matched resistive load to the rest of the line.

Alternate impedance-changing methods make use of a small ($\frac{1}{10}$ watt) resistor either in series with the inner conductor of the line or shunted between inner and outer conductor. A good impedance match on a 50-ohm line is possible with a resistor having a nominal value of 50 ohms. Sometimes it is possible to obtain a sufficiently good match by placing the resistor at the end of the line and making it serve as the coupling loop.

9. Balanced-to-unbalanced Transformers. As explained in Art. 1 of Chap. VII, a coaxial line is used to connect the transmitter and antenna

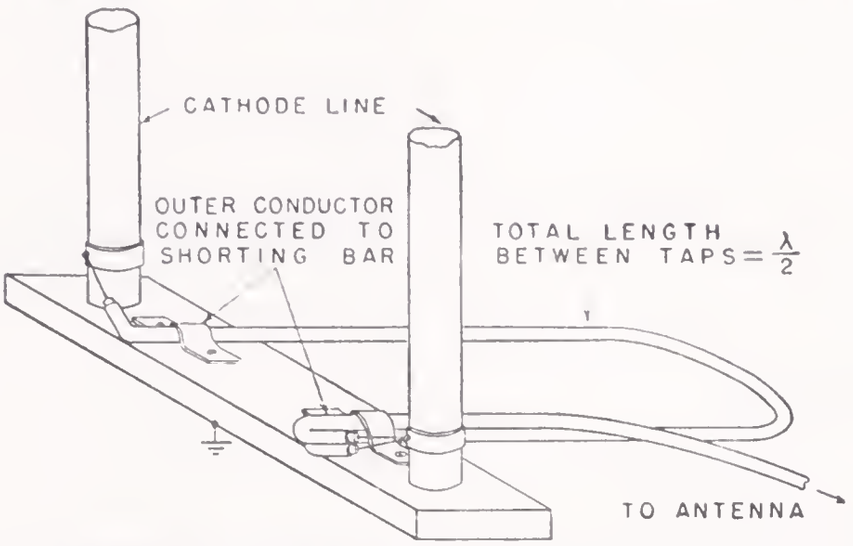


FIG. 33. Balanced-to-unbalanced transformer used with uhf oscillator.

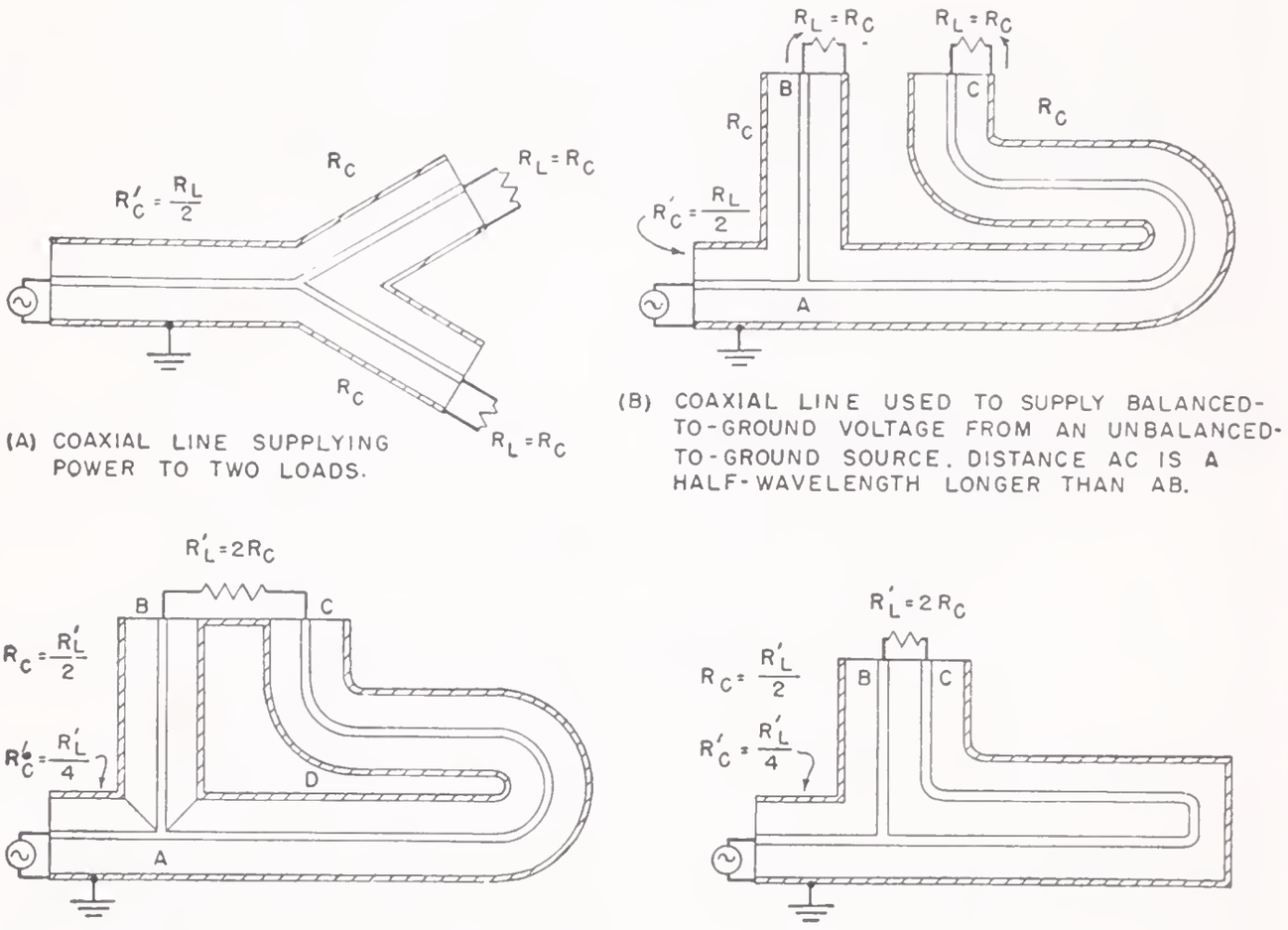
of nearly all uhf radar systems and many 10-cm microwave systems. Use of a parallel-wire line for this purpose is generally undesirable because radiation from the line causes an appreciable reduction in the power delivered to the antenna. Parallel-wire lines are very frequently used, however, as the resonant elements of uhf transmitters. Because these lines are usually balanced to ground, they cannot be connected directly to coaxial lines, the outer conductor of which is always grounded. Connection must be made through a *balanced-to-unbalanced transformer*.

A balanced-to-unbalanced transformer is also necessary at the antenna end of the coaxial line because radar antennas are normally composed of one or more pairs of radiating elements, each element having essentially the same impedance to ground. Attaching such an antenna to a coaxial line upsets this balanced-to-ground condition unless connection is made through an appropriate transformer.

The operation of a balanced-to-unbalanced transformer may be explained from Fig. 33, which shows the device used for connecting the cathode line of a uhf oscillator to a coaxial transmission line. The inner conductor of the antenna coaxial line is connected to one tap of the

cathode line, and the outer conductor is attached to the shorting bar. A coaxial line one-half wavelength long is connected between the two taps of the cathode line to avoid unbalanced loading of the oscillator.

Explanation of this device is most easily made in terms of transmission in the reverse direction, that is, from the unbalanced-to-ground coaxial line to a balanced-to-ground circuit. Actually, the device may be used equally well for transmission in either direction. In Fig. 34*A*, an r-f



(A) COAXIAL LINE SUPPLYING POWER TO TWO LOADS.

(B) COAXIAL LINE USED TO SUPPLY BALANCED-TO-GROUND VOLTAGE FROM AN UNBALANCED-TO-GROUND SOURCE. DISTANCE AC IS A HALF-WAVELENGTH LONGER THAN AB.

(C) BALANCED-TO-UNBALANCED TRANSFORMER SIMILAR TO THAT OF FIG. 33.

(D) ALTERNATE FORM OF BALANCED-TO UNBALANCED TRANSFORMER.

FIG. 34. Derivation of balanced-to-unbalanced transformer.

voltage source feeds a coaxial transmission line that connects to two other coaxial lines in parallel, each of which is terminated by a load resistance R_L equal to its characteristic resistance R_c. If the characteristic resistance of the first coaxial line is $R_c/2$, then no standing waves appear in the circuit, and the resistance presented to the source is $R_c/2$.

In Fig. 34*B*, the same line system is shown, except that the distance AC is carefully made one-half wavelength longer than the distance AB. Thus the voltage at the center conductor of end B is of opposite phase to the voltage at the center conductor of end C, and the voltages supplied to the two resistors R_c are balanced to ground. If current flows from the center conductor at end B to the resistor at a particular instant, then at the same instant current flows from the resistor into the center conductor

at end C, as indicated by the arrows. Conditions are not changed, therefore, if the outer conductors are shorted together at B and C and the two resistors are joined together, as shown in Fig. 34C. In this way, balanced-to-ground voltages are supplied to the load resistor $R_L' = 2R_c$ from a source having one end grounded.

The device in Fig. 34C is essentially the same as the balanced-to-unbalanced transformer shown in Fig. 33. If in Fig. 34C the distance AB is reduced to zero and the distance AC to a half wave, and if a portion of the outer conductor near point A is removed, the result is identical with the device in Fig. 33. Since the distance AC is one-half wavelength, the characteristic resistance of the section of line between A and C need not be R_c. In Fig. 34C a resistance $2R_c$ placed at the balanced-to-ground end of the transformer results in a resistance $R_c/2$ at the unbalanced-to-ground end Therefore, there is a transformation ratio of 4:1 for impedances in this device, and in Fig. 33 *the impedance placed between the two taps is four times the input impedance to the antenna transmission line.*

An alternate form of the transformer, shown in Fig. 34D, is obtained from Fig. 34C by collapsing the wall surrounding the space D so that it becomes a partition within the outer conductor. Removal of this partition results in the device of Fig. 34D.

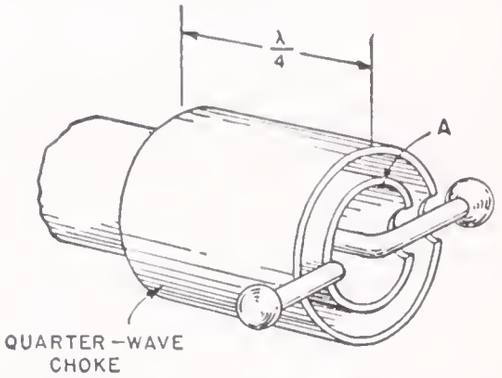


FIG. 35. Quarter-wave-choke type of balanced-to-unbalanced transformer.

In the microwave region, a common type of balanced-to-unbalanced transformer employs a quarter-wave choke (see Art. 1). In Fig. 35, a quarter-wave-choke transformer is used to connect an unbalanced-to-ground coaxial line to a center-fed half-wave antenna (a balanced-to-ground device). Because the input impedance of the choke is extremely high, the outer conductor of the coaxial line is effectively isolated from ground at point A and may therefore be connected to the antenna without disturbing the balanced-to-ground condition of the antenna. Observe that no impedance transformation is provided by the choke. The impedance presented to the coaxial line is equal to the input impedance of the balanced-to-ground device connected to the line.

10. Adjustment of Impedance Levels. For proper operation of r-f transmitting and receiving systems, the impedance at various points of the r-f line must be adjusted to certain levels during the transmission of pulses and the reception of echo signals. The required impedance levels are obtained by properly locating the t-r device in the branch line to the receiver and by use of impedance-changing devices at various points of the r-f transmission system.

The location of the t-r device in the receiver branch line depends upon the manner in which the lines are connected at the T junction. Basic methods of connection are indicated in the simplified diagrams of Fig. 36. For convenience, the transmission lines are drawn in these diagrams as parallel-wire lines; actually they may be any of the two-conductor types described in Chap. VII or any of the waveguide types discussed in Chap. VIII.

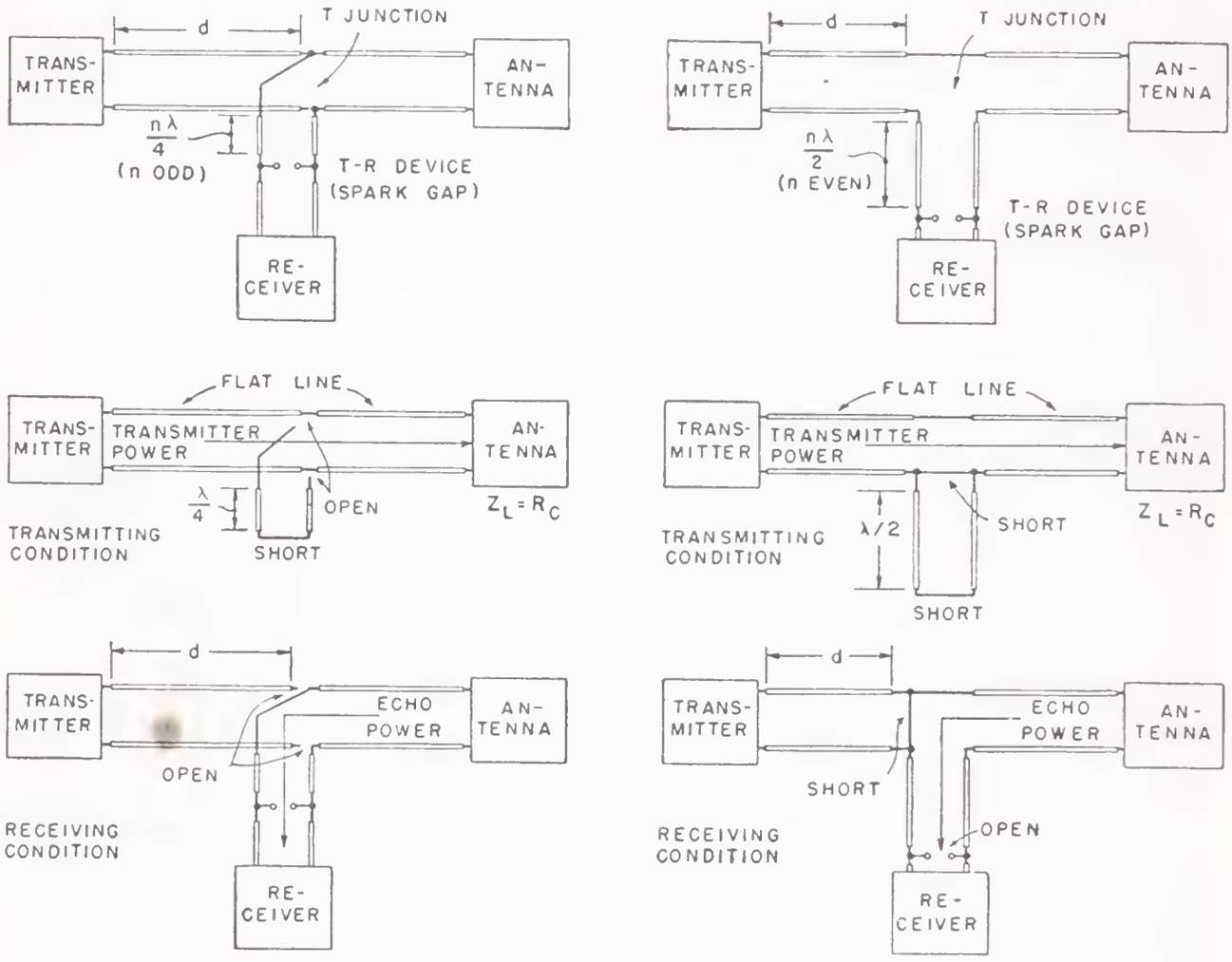


FIG. 36. Basic transmitting and receiving systems.

The t-r device represented in Fig. 36 is a spark gap which breaks down (ionizes) in the presence of the transmitting pulse and places a short circuit across the line at the gap location. If the parallel connection of lines is used, as in Fig. 36*A*, the t-r device must be located an odd multiple of quarter waves from the T junction. Locating the t-r gap in this manner causes the receiver branch line to appear as an open circuit across the transmitter line during the transmission of the pulses and consequently enables power to be transferred from transmitter to antenna without loss. A similar effect during transmission is secured in the series arrangement of lines, shown in Fig. 36*B*, by placing the t-r gap an even number of quarter waves from the T junction. The receiver branch

line then appears at the T junction as a short circuit in series with one of the transmitter-line conductors.

The distance d between transmitter and T junction in Fig. 36 is critical and depends upon whether the series or parallel arrangement of lines is employed. Between pulses the transmitter has, in general, an impedance greatly different from the characteristic impedance of the line connected to it. During the reception of echo signals, therefore, a high standing-wave ratio appears on the line section between transmitter and T junction. For the parallel connections of lines, d must be chosen so that the transmitter line presents maximum impedance to the echo signals at the T junction; for the series arrangement of lines, d must be selected for minimum impedance in the direction of the transmitter at the T junction.

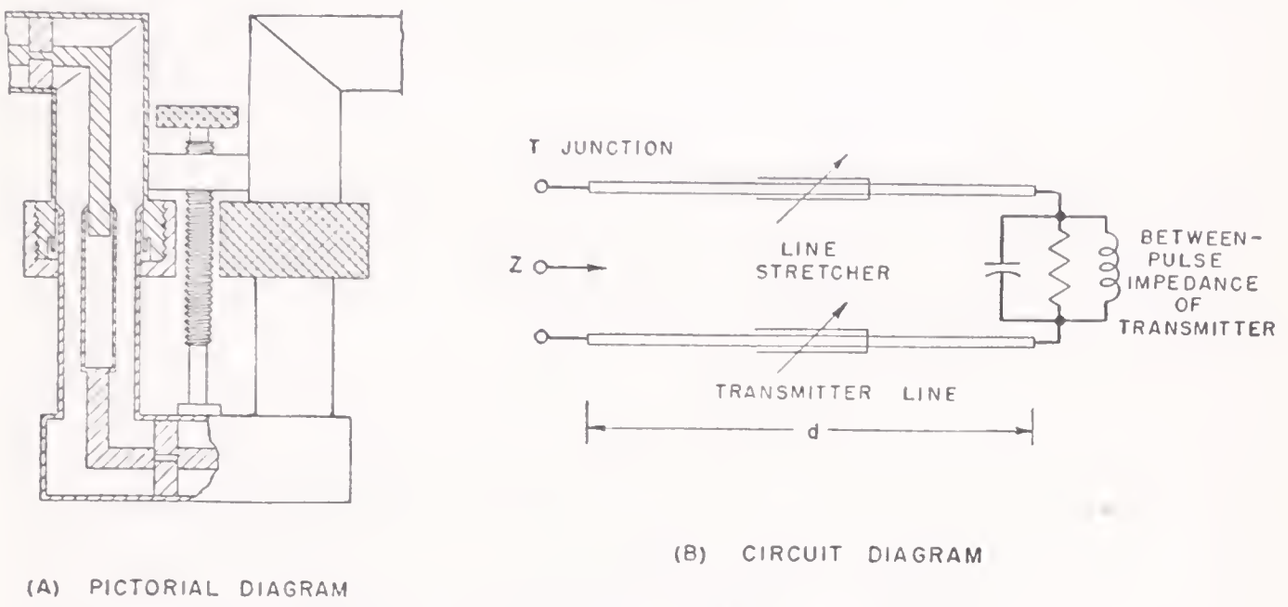


FIG. 37. Line stretcher for use in coaxial lines.

Selecting d in accordance with these principles has the effect of disconnecting the transmitter from the system between pulses and of enabling the echo signals to be transmitted past the unfired t-r gap to the receiver.

The r-f systems of Fig. 36 have the disadvantage that the distance d depends upon the transmitter impedance between pulses. Changes in this impedance brought about by changes in the transmitter cause a shift in the standing-wave pattern on the transmitter line. Consequently, d must be readjusted to provide optimum receiver conditions whenever changes are made in the transmitter. Furthermore, because the transmitter impedance contains a small resistive component, the standing-wave ratio on the line is less than infinity. As a result, some echo power is lost to the transmitter even when d is correctly adjusted.

In some microwave systems, d is adjusted for proper operation with a specific type of magnetron during manufacture. Normally, magnetron characteristics are sufficiently uniform so that no further adjustment is required when a magnetron is replaced. In systems employing parallel-

wire or coaxial lines, an adjustable length section is sometimes included in the line between transmitter and T junction. This section, called a *line stretcher*, is represented pictorially in Fig. 37*A* and schematically in Fig. 37*B*. The line stretcher enables d to be adjusted so that, between transmitted pulses, the input impedance at the T junction is maximum for the parallel arrangement of the lines and minimum for the series arrangement. The adjustment is usually made experimentally by observing a target on

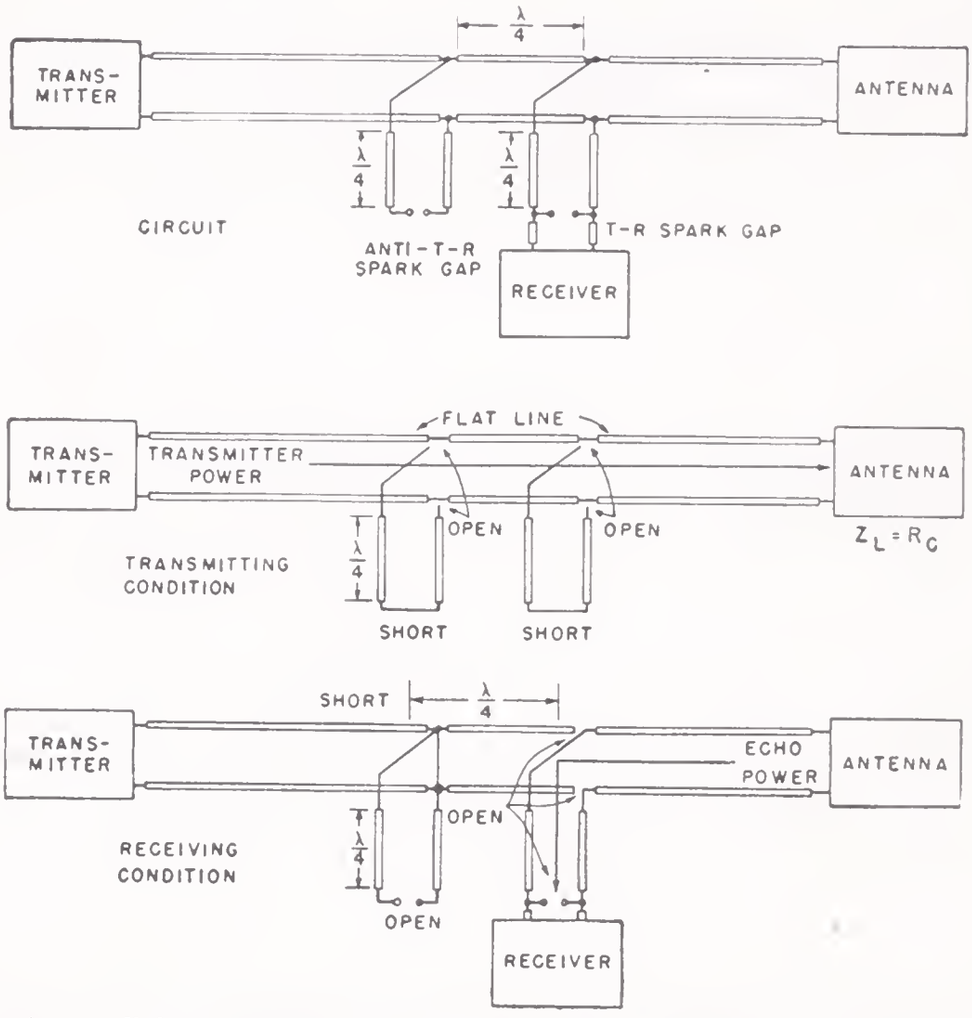


FIG. 38. Basic parallel transmitting and receiving system using t-r and anti-t-r spark gaps

the radar indicator and setting the line stretcher for maximum signal strength.

An alternative method of minimizing the effect of transmitter impedance on echo signals makes use of an anti-transmit-receive (anti-t-r) device in the line section between transmitter and T junction. The anti-t-r device, like the t-r device, may be represented as a spark gap terminating a quarter-wave section of the transmission line (see Fig. 38). During transmission, both the t-r and anti-t-r gaps are fired, and each produces a short circuit at the gap and an open circuit a quarter wave away at the main transmission line. During reception, neither gap is fired, and because the anti-t-r gap is a quarter wave from the main line, a short circuit is produced at its T junction. If this short circuit is a quarter wave from the receiver branch line, as in Fig. 38, the transmitter

is effectively disconnected and echo signals are directed toward the receiver. If series connections at the T junctions are used, operation is the same except that the distances from the spark gaps to the main line must be made an even multiple of half wavelengths. Use of an anti-t-r device eliminates the need for a line stretcher because changes in the between-pulse transmitter impedance affect only the impedance in parallel with the short circuit at the anti-t-r T junction. The transmitter impedance can have no effect upon echo signals, and, furthermore, a resistive component of transmitter impedance causes no loss of echo power.

If the t-r device in Fig. 36 is operating properly, the branch line to the receiver may be ignored during the transmission of pulses and the r-f system may be represented by a single transmission line joining trans-

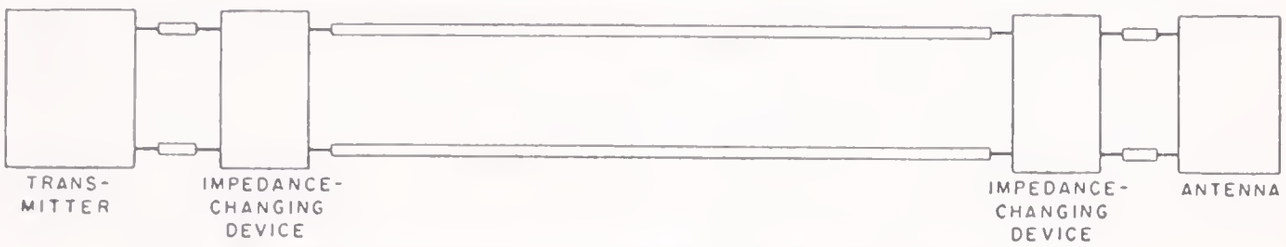


FIG. 39. Basic form of the r-f system during transmission of pulses.

mitter and antenna (see Fig. 39). Two impedance-changing devices are normally required on this line—one near the antenna and another near the transmitter. The antenna impedance-changing device serves to match the antenna impedance to the characteristic resistance of the line, thus enabling the line to operate with unity standing-wave ratio. Flat-line operation is desirable because the line then has maximum power capacity and highest efficiency and is least sensitive to small changes in frequency. Absence of standing waves on the transmitter line also allows a line stretcher to be adjusted for maximum echo signal without upsetting the impedance presented to the transmitter.

A double-stub transformer (Art. 15, Chap. VII) is sometimes used to match the antenna to a coaxial line, but more frequently a multiple quarter-wave transformer in the form of a tapered line section is employed. The tapered section provides a good impedance match over a broad band of frequencies and requires no adjustment after installation. In waveguide r-f systems, a single iris or two irises spaced approximately three-eighths wavelength apart may serve as the impedance-changing device. Inductive irises are usually employed because of their ability to withstand large field intensities without danger of arc-over. On parallel-wire lines, the most commonly used matching device is a single stub adjustable in length and position on the line.

An impedance-changing device is required near the transmitter because a load impedance of R_c on a radar transmitter does not, in general, result

in optimum performance of the radar system. For optimum performance, the transmitter should have high power output and good frequency stability. Increasing the power output improves the range of a radar set only if its operating frequency remains well within the bandwidth of the receiver. The receiver can be tuned to the transmitter frequency only if the transmitter frequency is constant or (when automatic frequency control is employed) if the transmitter frequency shifts gradually between relatively narrow limits. As discussed in Art. 10 of Chap. X and Art. 10 of Chap. XI, uhf and microwave transmitters are likely to be least stable when delivering maximum power output. The final adjustment of transmitter load impedance, which is the input impedance of the r-f line, is therefore a compromise between high power output and good frequency stability. The proper value of transmitter load impedance for best over-

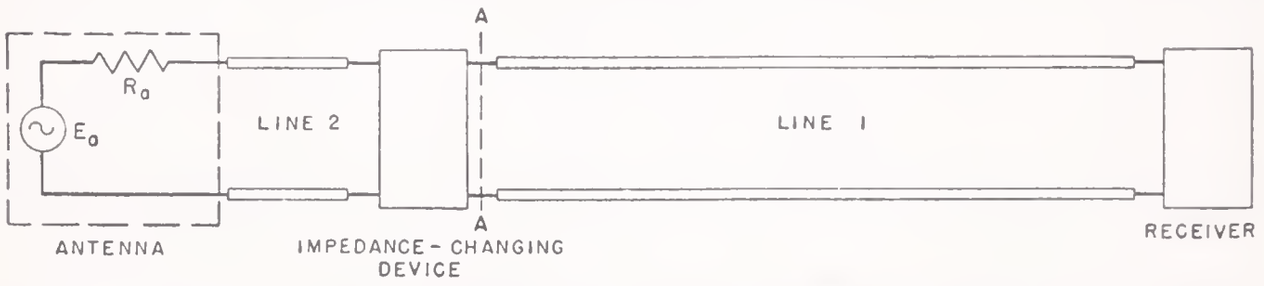


FIG. 40. Basic form of the r-f transmission system during the reception of echo signals.

all system performance may be determined from a Rieke diagram of the transmitter (Art. 10, Chap. XI). Use of an impedance-changing transformer near the transmitter permits attainment of the desired load impedance.

The impedance-changing device at the transmitter may be similar to any of those used at the antenna end of the line. Very often in magnetrons, the device is made an integral part of the output termination (see, for example, the magnetron of Fig. 21*B*, p. 35). When an impedance transformer is not included in the magnetron, the impedance-changing device is located near the magnetron in the r-f line. With coaxial lines, a quarter-wave transformer is normally used (see Fig. 3); in waveguide systems, an inductive iris is commonly employed.

Figure 40 shows the basic form of the r-f transmission system during the reception of echo signals. This circuit assumes that the line section between T junction and transmitter has no effect upon echo signals. In the diagram the generator with voltage E_a accounts for the voltage induced in the antenna by the echo signal and R_a represents the impedance of the antenna. If half-wave elements are assumed, the antenna impedance may be considered to be resistive and to have essentially the same value during transmission and reception. The impedance-changing device near the antenna serves to terminate line 1 in a matched load during the transmission of pulses.

If the input impedance of line 1 at point A is R_c in the direction of the receiver, a conjugate impedance match will exist at point A and maximum echo power will be transferred to line 1. If line 1 is lossless, this adjustment also ensures maximum power to the receiver input terminals. Thus the receiver input circuit should terminate line 1 in R_c for maximum signal strength. Actually because maximum signal-to-noise ratio rather than maximum signal alone is desired, the receiver input impedance is made slightly different from R_c. The method of adjusting for maximum signal-to-noise ratio depends upon the type of input circuit employed. Several types of receiver input circuits and details of their adjustment are described in Arts. 15 to 17 of this chapter.

11. Tubes for T-R Devices. Unfortunately a spark gap is not the ideal switch assumed in the preceding article. Rather, it behaves as a high impedance when not fired and as a nonlinear resistance that maintains a nearly constant low voltage across itself when fired. Furthermore, its characteristics change somewhat with use. Expedients used to increase its effectiveness as a t-r or anti-t-r device include enclosing it in a partially evacuated glass envelope and making the device part of a one- or two-entry resonator. Figure 41 illustrates a typical t-r tube for use in coaxial-line uhf transmission systems. Microwave t-r tubes may be seen in the t-r assemblies of Figs. 48 to 52. Characteristics of t-r tubes are presented in this article.

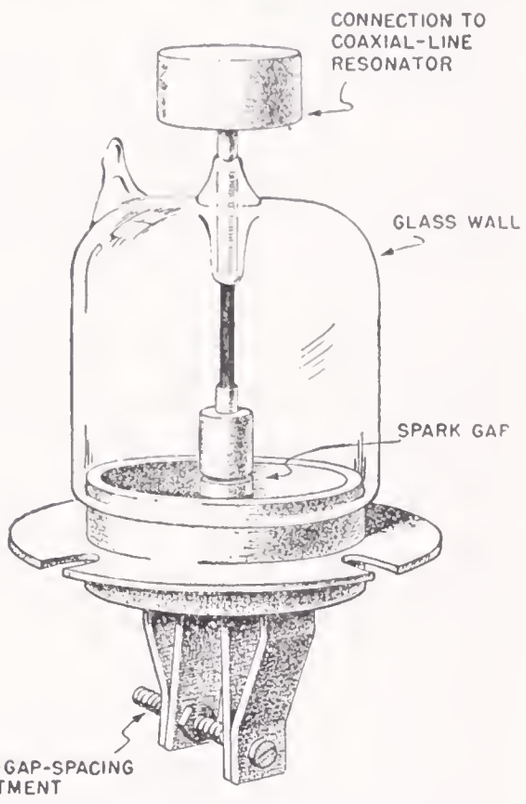


FIG. 41. Typical uhf spark-gap tube.

Enclosing the spark gap in a glass envelope and reducing the pressure of the atmosphere around the gap to a few millimeters of mercury have the effect of reducing both the breakdown and running voltages of the gap. The envelope of the r-f voltage across the spark gap in an evacuated t-r tube during a transmitter pulse varies in the manner indicated by Fig. 42. The "spike" at the front of the pulse results because a relatively high voltage is required to fire the gap. The lower voltage occurring after the spike is the voltage maintained by the discharge after breakdown. Both the breakdown and the running voltages depend mainly upon the nature and pressure of the gas in the t-r tube. These voltages must be made as small as possible to provide good receiver protection.

In microwave systems, a keep-alive electrode is provided to reduce the spike to as small an amplitude and duration as possible. The keep-alive is an auxiliary electrode mounted near one of the gap terminals. A negative direct voltage is applied to it so as to maintain a steady glow discharge near the t-r gap. Within the glow discharge, electrons move from the keep-alive electrode to the gap electrode, and a few of them enter the t-r gap when radio-frequency voltage is applied. Because of these electrons, the gap is fired more quickly.

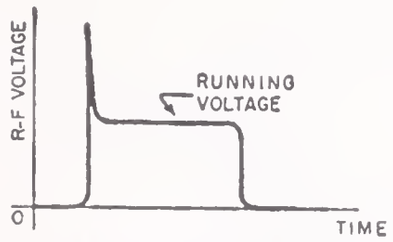


FIG. 42. Envelope of r-f voltage across gap in t-r tube.

When a radar system is turned off, no keep-alive voltage is available. To avoid the possibility of strong pulses from near-by radar systems damaging the crystal, most microwave systems are provided with *crystal shutters* (see Art. 16).

Even when every effort is made to keep the voltage at the t-r tube low during the transmitted pulse, the value obtained is much higher than the allowable input voltage for receivers using crystals. A secondary difficulty is the relatively large part of the transmitted power wasted in the t-r discharge. The effectiveness of the t-r device can be improved by using the equivalent of an ideal transformer, as shown in Fig. 43*A*. The transformer steps up the main-pulse voltage to produce a much higher voltage across the spark gap on the secondary side, with the result that the breakdown voltage across the gap corresponds to a lower voltage across the transmission line. The voltage across the line during the discharge is reduced also. The voltage waveform at the receiver is of the

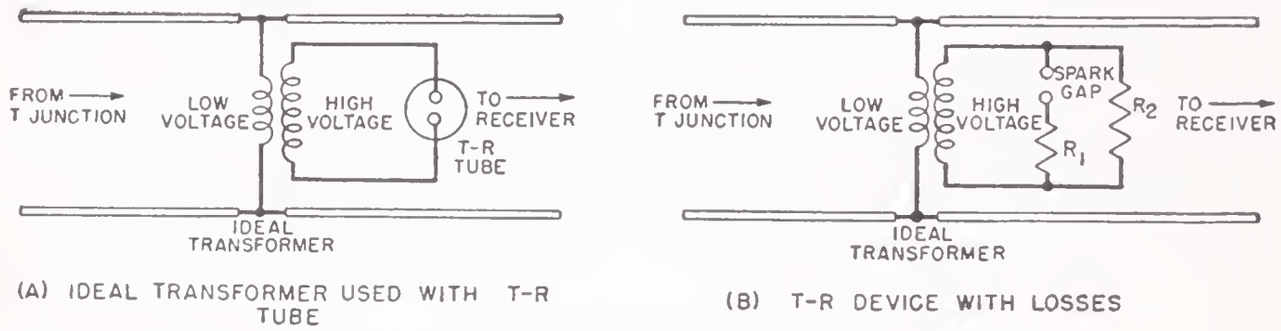


FIG. 43. Equivalent circuits of t-r devices.

same shape as that across the t-r tube (see Fig. 42), but the voltage at each instant is reduced by the transformation ratio, and receiver protection is improved. Furthermore, less power is absorbed from the transmitter because the lower voltage at the t-r position on the transmission line means a closer approach to a short circuit there. Because of the high frequencies used in radar, the transformer of Fig. 43*A* actually takes the form of a one- or two-entry resonator. The particular devices used in uhf and microwave systems are discussed in Arts. 12 to 14. Although these devices cannot be considered even approximately as ideal trans-

formers when the spark gap is fired, they do, nevertheless, materially reduce the power consumed by the spark gap during transmission.

The circuit of Fig. 43*A* would seem to indicate that the transformation ratio should be as high as possible in order that the t-r tube may provide maximum protection. An actual circuit is represented more nearly by Fig. 43*B*, in which R_1 represents the nonlinear resistance of the gap when fired and R_2 is a very high resistance representing the losses of the circuit. A very high value of transformation ratio causes R_2 to appear as a relatively low resistance across the receiver line during reception. Thus the choice of transformation ratio must be a compromise between a high value for receiver protection and a low value for low attenuation of received signals.

After a glow discharge has been produced in a gas and the voltage which produced that discharge has been removed, large numbers of free electrons and ions remain in the gas for a substantial time. The electrons and ions slowly recombine or migrate to electrodes in the surrounding structure. Such ions are present in a t-r tube following the transmitted pulse. Voltages resulting from received signals appear across the spark gap and set the ions in motion. A portion of the received power is wasted in imparting motion to the ions. The t-r structure therefore causes attenuation of the received signal until recombination of all ions has been effected. The recovery time is defined as the interval between the end of the transmitted pulse and the time at which the power in the signal applied to the receiver is one-half as large as it would be if no ions were present.

From the viewpoint of recovery time alone, *water vapor* is the most effective stable gas for use in t-r tubes. It has shortcomings, however, the principal one being that the pressure varies appreciably with ambient temperature so that performance is erratic. Temperature-controlled t-r devices have been used, but the complication is in general unwarranted.

A noble gas, such as argon, is desirable from the point of view of low discharge voltage and good receiver protection. The recovery time of noble gases, however, is so long that they are not generally used, except in long-range early-warning equipment where short minimum range is unimportant.

The most common gas filling for short-range t-r tubes is *air*, and often air in which no precaution has been taken to remove water vapor. In the aging process to which the tube is subjected in manufacture, certain components of the air filling are eliminated. The resulting gas content is such as to produce a recovery time of about 1 or 2 μs when the tube is first put into service.

The effectiveness of a t-r tube in protecting a receiver may decrease with age of the tube because of a gradual decrease of gas pressure that occurs. The gas is gradually consumed by chemical combination with

electrodes and by the gas molecules becoming embedded in the gap electrodes under the influence of the glow discharge, especially the keep-alive discharge. To prevent unnecessarily rapid loss of gas, the keep-alive current should be kept as low as is consistent with receiver protection. Electrodes in spark-gap tubes are sometimes pre-oxidized to minimize the chemical combination.

The effect of gas pressure is shown by the curve of Fig. 44. The tube originally is filled to a relatively high pressure, of the order of 7 to 10 mm of mercury, in order to obtain adequate life. This pressure is customarily chosen to afford the receiver the minimum permissible amount of protection. As the tube is used, the gas is consumed gradually, and an improvement in receiver protection is obtained until the pressure falls to the minimum-voltage point of the curve. Still lower pressures result in higher gap voltages and less receiver protection. The protection gets less very rapidly as the pressure decreases. If the t-r tube is used for a very long time, a point is eventually reached at which the transmitter cannot fire the gap at all. However, the tube is worthless for receiver protection considerably before this point is reached. The tube must be replaced before the final end point is actually reached. Often a t-r tube that is nearing the end of its life is used in the anti-t-r device, as poor performance in that position cannot damage the receiver.

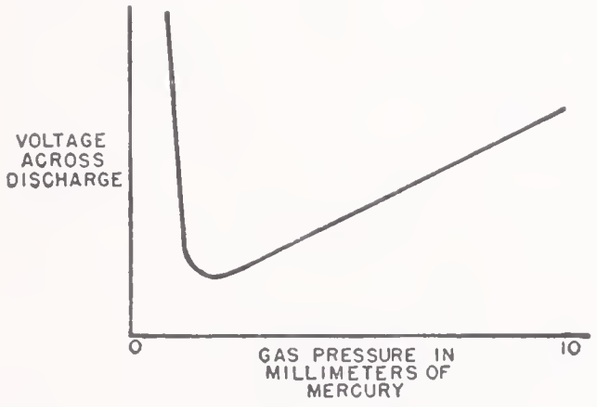


FIG. 44. Variation of running voltage of discharge with gas pressure in a t-r tube.

Sometimes t-r tubes become useless with age for a different reason. The process by which gas is absorbed from the tube is selective. Water vapor may be absorbed more rapidly than the other gases, and as a result, the recovery time is increased. Where minimum range is important, the reason for replacing t-r tubes is often excessive recovery time.

Still another effect occasionally limits the life of a t-r tube. In normal operation of the glow discharge, ions strike the metal electrodes (usually copper) with considerable force, frequently driving out molecules of the metal. These molecules may deposit themselves upon the inner wall of the glass envelope. This process is called *sputtering* and builds up a film of copper on the glass. The film causes losses in the t-r device and reduces the echo signal supplied to the receiver. Sputtering does not, however, reduce the protection afforded by the t-r tube.

12. Ultrahigh-frequency T-R Systems. Single-entry parallel-wire or coaxial-line resonators are commonly used as transformers in transmit-receive devices operating at frequencies from 100 to about 700 Mcps.

The important features of a simple parallel-T-junction transmit-receive switch employing a parallel-wire-line resonator are illustrated in Fig. 45*A*.

A section of transmission line slightly less than a quarter wavelength long, short-circuited at one end and open-circuited at the other end except for the presence of a spark gap, is connected across the receiver branch of the transmission line at a point approximately a quarter wavelength toward the receiver from the T junction. Sometimes t-r tubes are employed in systems of this kind, but because crystals are not generally used in uhf receiver-input stages, only moderate protection is required, and an unenclosed air gap is often sufficient. During reception, the shorted-end section of line resonates with the capacitance of the unfired

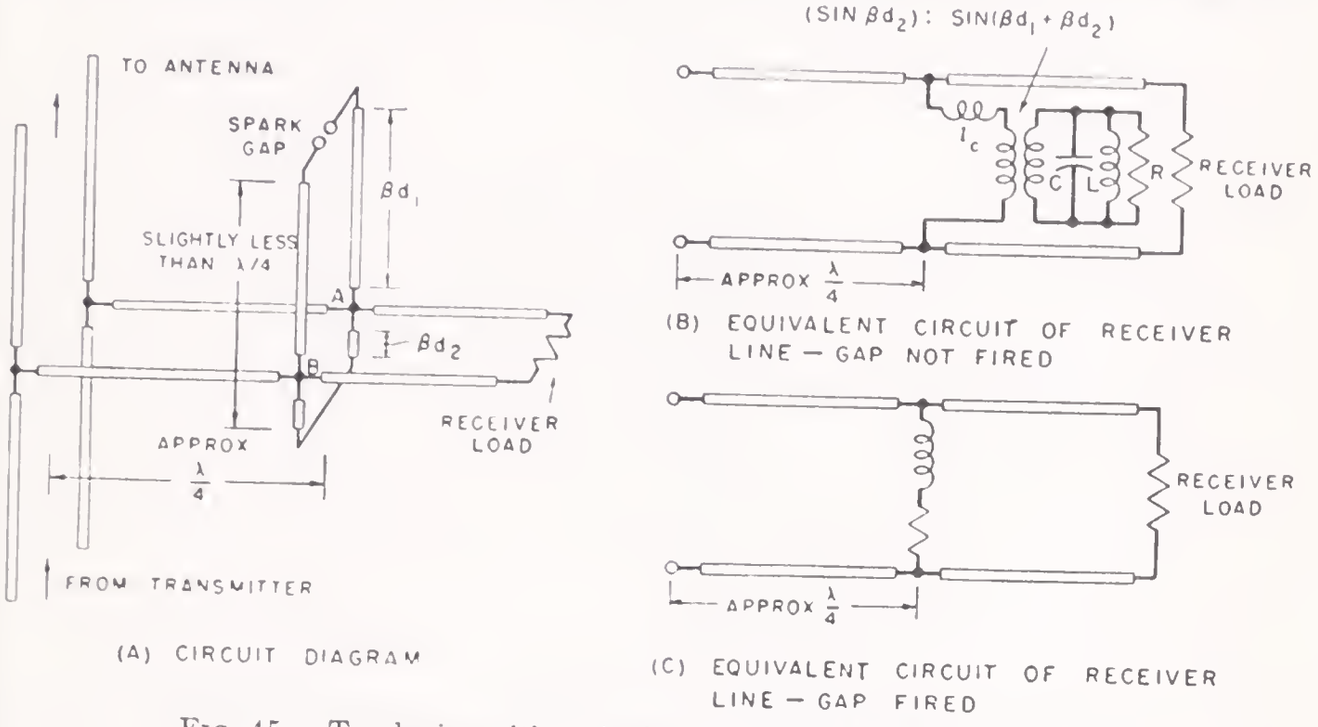


FIG. 45. T-r device with a single-entry resonator and spark gap.

spark gap and, in the lowest open-circuit mode of resonance, presents a very high impedance to the receiver branch line. This fact may be seen from an equivalent circuit of the device drawn in Fig. 45*B* in accordance with the principles of Arts. 8 and 9, Chap. IX. Inductance l_c in the equivalent circuit represents the coupling inductance, and the ideal transformer accounts for the effect of the tapped connection on the resonator, represented in Fig. 45*B* by R, L, and C in parallel. At resonance the impedance across the receiver line is the coupling reactance in series with the very high impedance of the resonant circuit. Thus the resonant line has negligible effect upon echo signals.

When the transmitter is energized and the spark gap breaks down, the capacitance of the unfired gap in Fig. 45*B* is replaced by the low resistance of the ionized gap. The t-r load therefore becomes essentially the input impedance of the βd_2 length of line across the receiver branch line (see Fig. 45*C*). The input impedance of the βd_2 section may be deter-

mined graphically from the Smith chart. If d_2 is made small by locating the taps near the short-circuited end of the resonator, the input impedance can be made small compared with the characteristic resistance of the receiver line. Thus a high impedance can be made to appear at the T junction in the direction of the receiver by locating the t-r resonant transformer slightly less than a quarter wave from the T junction.

An anti-t-r device can be used with the open-wire transmission-line system of Fig. 45. The quarter-wave resonator of such a device is identical with the t-r resonator except that no output terminals are provided for connection to a receiver.

Ultrahigh-frequency radar systems operating above about 200 Mcps use coaxial rather than open-wire transmission lines, and coaxial resonant

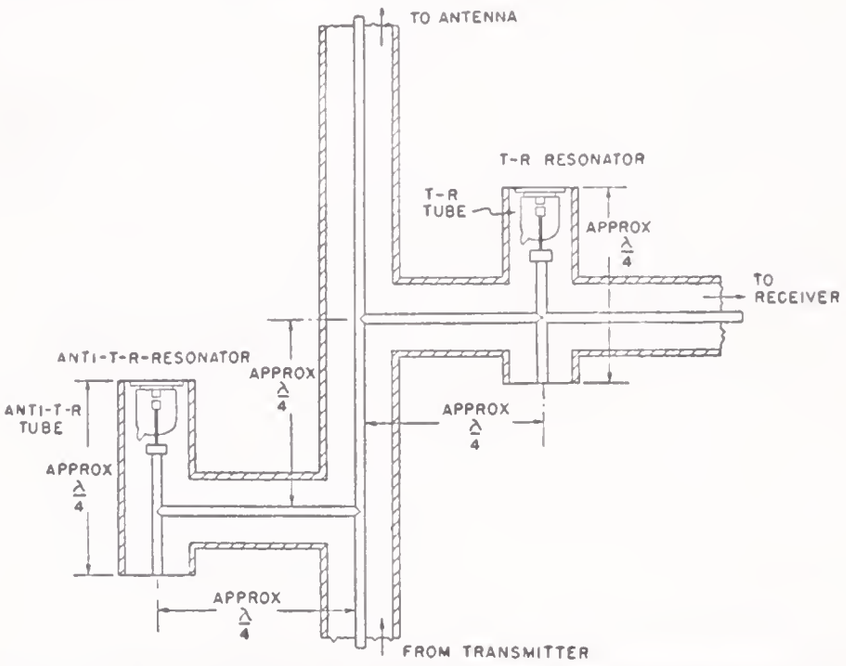


FIG. 46. Transmit-receive device using a coaxial quarter-wavelength single-entry resonator for the t-r and anti-t-r switches.

transformers are used in the transmit-receive devices of these systems. Coaxial quarter-wave resonators are identical in principle with open-wire quarter-wave resonators. Figure 46 shows the physical construction of both t-r and anti-t-r coaxial-line resonators. The lower end of each of the vertical quarter-wavelength line sections is short-circuited. At the upper end of the line, the center conductor is connected to a t-r tube of the type illustrated in Fig. 41. The adjustment shown beneath the tube in Fig. 41 is used to obtain correct gap spacing, thus compensating for wearing down of the gap electrodes. When the gap spacing is changed, the coaxial-line resonator must be retuned, since the capacitance of the gap is changed.

The resonant line for the t-r devices need not be of the single-entry type. For example, Fig. 47 shows a two-entry coaxial-line resonator which is approximately a half wave in length and is short-circuited at each end. The standing-wave pattern of voltage between conductors in the resonator has a zero point at each end and a maximum point at the center

where the t-r gap is located. Input and output connections to this resonator are made near opposite ends of the half-wave line, the distances being chosen to provide correct transformation ratios.

13. Transmit-Receive Devices for Coaxial-line Microwave Systems. At frequencies of 3,000 Mcps and higher, resonant cavities are of reasonable size and are used in place of resonant-line sections in t-r devices because of their smaller losses. Because the physical dimensions of a practical microwave t-r tube (such as that shown in Fig. 48*A*) are comparable to a wavelength, the tube is designed as part of the cavity.

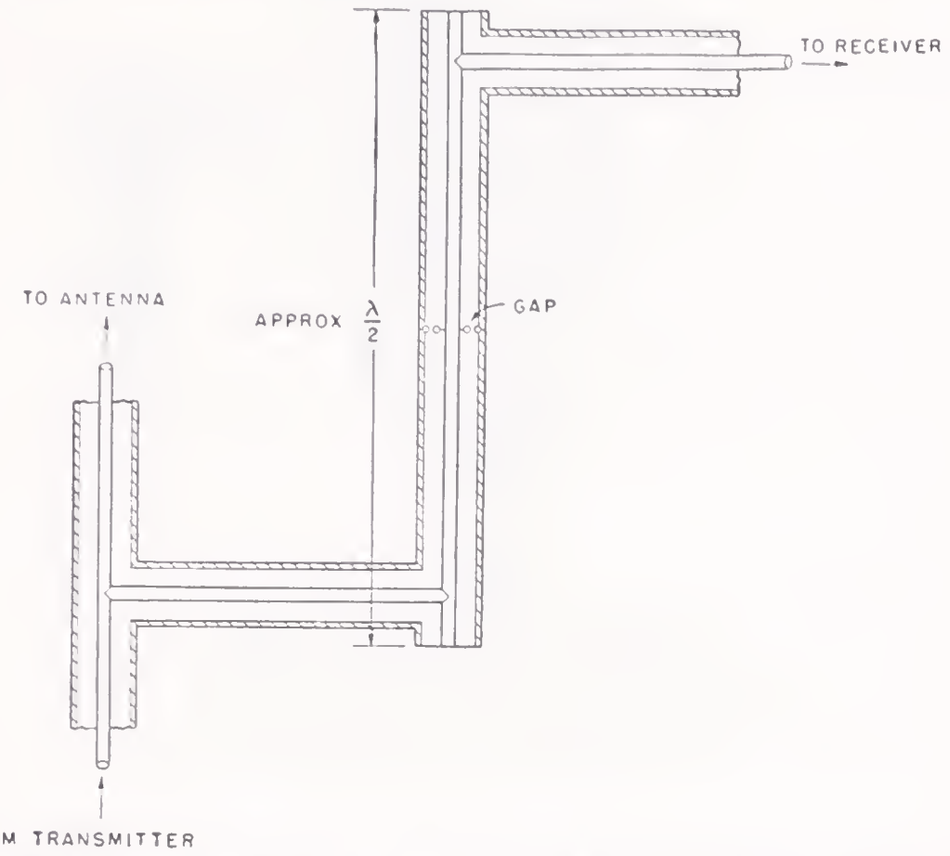
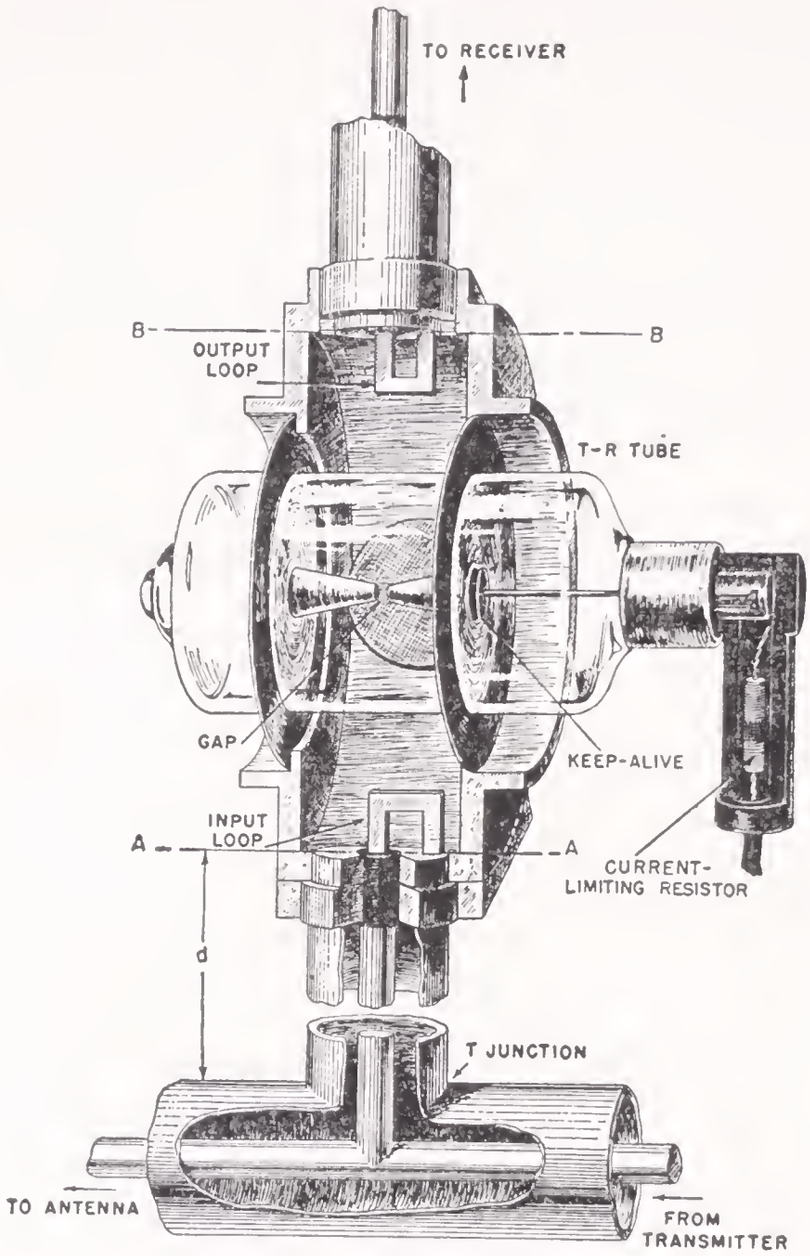


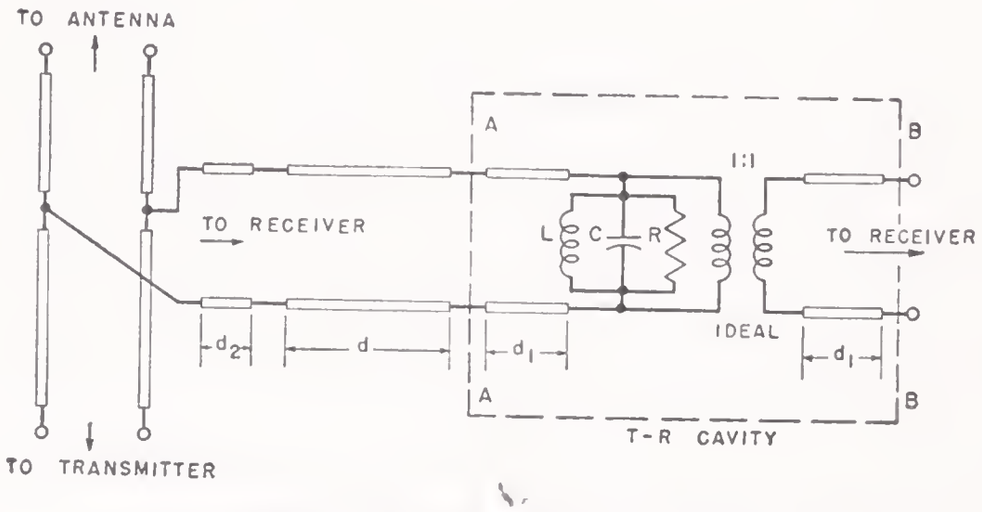
FIG. 47. Transmit-receive device using a two-entry half-wavelength coaxial-line resonator.

The t-r device of Fig. 48*A* is typical of those used in medium-power 10-cm radar systems employing coaxial lines. The gap is produced by the two funnel-shaped, hollow center posts approaching each other inside the vacuum tube. Copper disks are sealed through the glass walls of the tube and connect to the center posts. The disks fit snugly into an external cylindrical section to form a two-entry re-entrant type of cavity. Coupling loops are used to connect the cavity to the input and output transmission lines, and a parallel type of connection is used at the T junction of transmitter and receiver lines.

The impedances at various points of the receiver branch line are indicated in Fig. 48*B*, which is an equivalent circuit of the receiver line between T junction and output terminals of the t-r device. The parallel combination of R, L, and C oscillates at a frequency corresponding to the open-circuit mode of the cavity, and the two line sections of lengths d_1



(A) PICTORIAL REPRESENTATION



(B) EQUIVALENT CIRCUIT EMPHASIZING OPEN-CIRCUIT RESONANT MODE OF CAVITY

FIG. 48. Microwave t-r device for systems employing coaxial lines.

account for the coupling inductances of the input and output coupling loops (see Art. 11 of Chap. IX). Coupling through the cavity is accounted for by the ideal transformer, a turns ratio of 1:1 being used because of the symmetry of the cavity. The coaxial-line section between T junction and point A is represented by parallel-wire lines of lengths d and d_2, the d_2 length being included in order to account for the coupling inductance at the T junction (see Art. 10 of Chap. IX).

Because the power in the echo signals is insufficient to cause breakdown of the t-r spark gap, oscillations build up in the cavity between transmitted pulses, the mode of oscillation being such that nearly all the input power to the cavity is coupled to the output transmission line. During the reception of echo signals, therefore, the high resistance R in Fig. 48B appears across the receiver line at the t-r-device location, and echo signals are propagated through the t-r device with negligible loss of power. During a transmitted pulse, oscillations in the mode that couples power from input to output terminals of the cavity are not possible because of breakdown of the spark gap. The nonresonant-cavity impedance therefore places an extremely low impedance across the receiver branch line. By making the distance $d + d_1 + d_2$ nearly a quarter wave, an extremely high impedance can be made to appear in parallel with the transmitter line during a transmitted pulse.

The t-r cavity in Fig. 48A is tuned by means of screw plugs which are inserted radially into the outer ring of the cavity. An alternative method of tuning uses a variable gap spacing, obtained by means of a specially constructed spark-gap tube. Changing the gap spacing changes the capacitance at the center of the cavity, thus changing the resonant frequency.

As shown in Fig. 48A, the connector for applying keep-alive voltage to the tube contains a resistor which is in series with the keep-alive circuit. A current-limiting resistor is necessary because of the nearly constant-voltage characteristic of the glow discharge. Often part of the resistance is located in the power supply, but at least some of it is always placed next to the t-r tube in order to prevent tube, resistance, and wiring capacitance from operating as a saw-tooth relaxation oscillator. If such oscillations should occur, the crystal in the receiver input circuit would be permanently damaged by transmitted pulses during the time the keep-alive gap is not fired.

Some 10-cm, coaxial-line, r-f systems employ apertures rather than coupling loops to connect the t-r cavity to the input and output transmission lines. An example of such a system is pictured in Fig. 49A. The t-r device is located between transmitter and receiver lines, the t-r tube being oriented so that the axis of its electrodes is parallel to both lines. Each line is coupled to the cavity by means of a small aperture. Since

this coupling arrangement effectively places the cavity in series with each line (see Art. 18, Chap. IX), the r-f system is equivalent to the series-T-junction type illustrated in Fig. 36*B* of this chapter.

For proper operation of the r-f system, the cavity is tuned to its open-circuit mode of resonance during the reception of echo signals, and through use of an anti-t-r device or an appropriate length of line, essentially a short circuit is made to appear at the T junction in the direction of the transmitter. The corresponding equivalent circuit with coupling reactances replaced by equivalent lengths of transmission lines is drawn in Fig. 49*B*. Between transmitter pulses echo signals are able to pass to the receiver with negligible loss of power. During transmission, the t-r

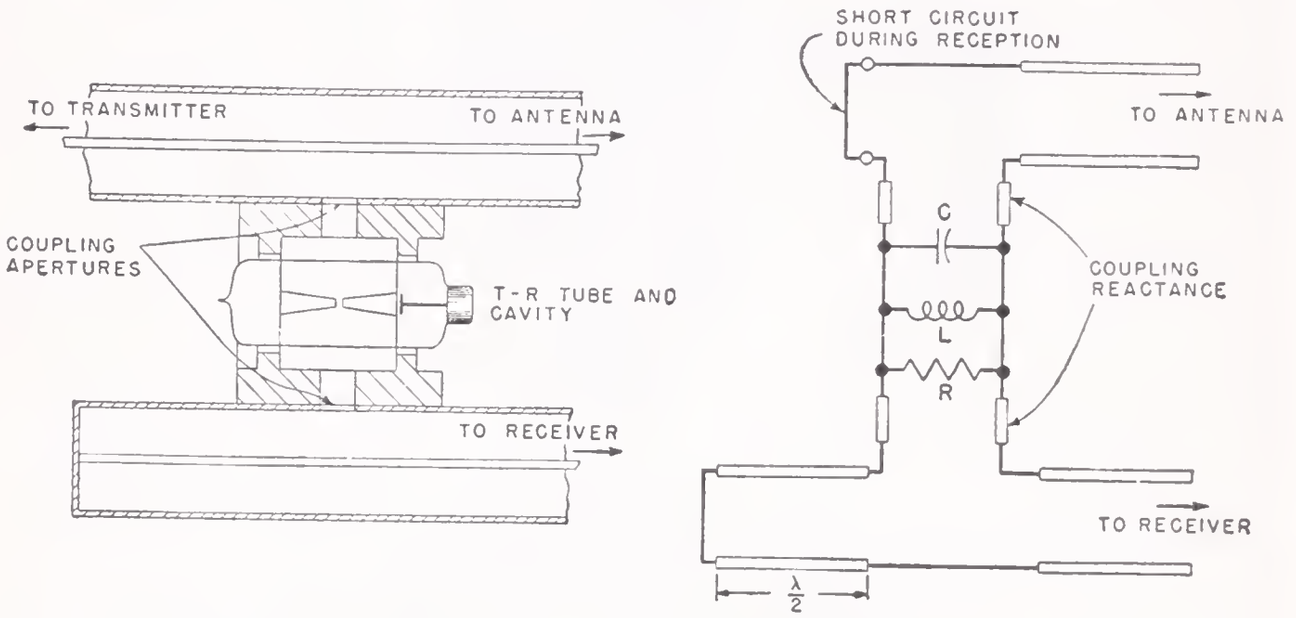


FIG. 49. Microwave t-r device aperture-coupled to transmitter and receiver lines.

gap is fired, and a very low impedance appears across the aperture in the transmitter line.

14. Transmit-Receive Devices for Waveguide Systems. The cavity-resonator and spark-gap-tube combinations illustrated in Figs. 48*A* and 49*A* can also be used for transmit-receive switching in waveguide r-f systems. Figure 50*A* shows a typical waveguide t-r system in which the t-r and anti-t-r cavities are aperture coupled to a narrow wall of the guide. The t-r and anti-t-r tubes, together with the supporting metal walls, form a re-entrant type of cavity similar to the cavity of Fig. 48*A*. The overall device is suitable for use in low- and medium-power radar systems operating at 10 cm.

A two-wire equivalent of the device is given in Fig. 50*B*. The cavities are represented in the usual manner by the lumped-constant elements R, L, and C and by transmission-line sections to account for coupling reactances of the apertures. The cavities are shown as terminations for quarter-wave line sections in parallel with the main line. The parallel

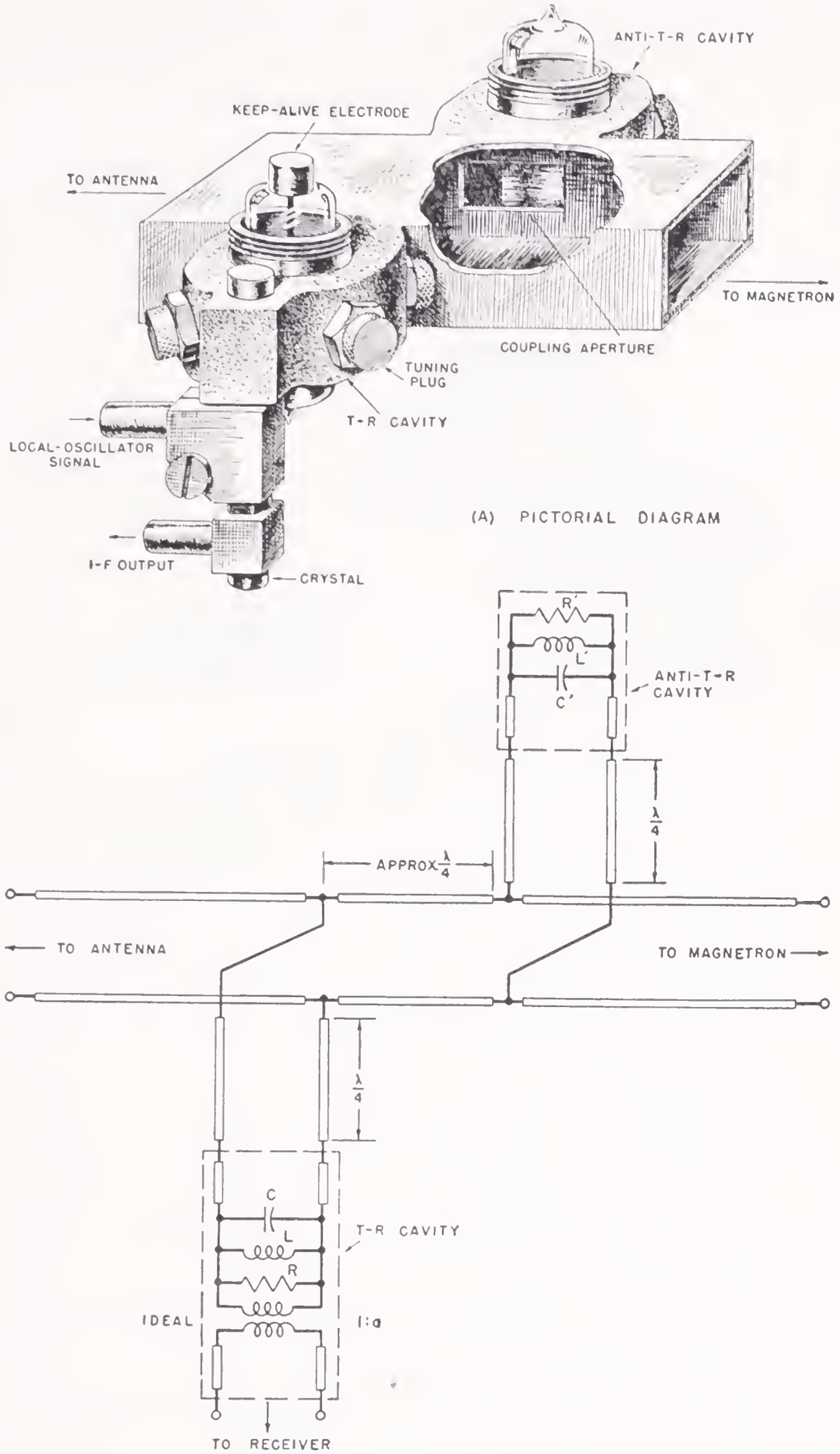


FIG. 50. Typical t-r device of a low- and medium-power waveguide system.

connections at the T junctions result from the fact that both cavities couple to a narrow wall of the main guide and hence form $\mathcal{H}$-plane junctions at the cavity locations (see Art. 5 of this chapter and Art. 18 of Chap. IX).

Operation of the circuit of Fig. 50*A* is similar to that of the basic parallel type of t-r and anti-t-r system illustrated in Fig. 38. During a transmitted pulse, both gaps fire and cause the cavities to appear as very low-impedance terminations for the quarter-wave line sections. Firing the gaps is thus very nearly equivalent to closing the side walls of the main guide, and as a result, the transmitter pulse passes directly to the antenna. During the reception of echo signals, both cavities are tuned to resonate in their open-circuit mode. Because the anti-t-r cavity is effectively a half wavelength from the T junction of the receiver branch line, the transmitter is in effect disconnected from the r-f system during reception. The t-r cavity appears as a high impedance in parallel with the branch line to the receiver, and echo signals are thus able to pass directly to the receiver.

In the t-r system of Fig. 50*A*, some power at the magnetron frequency passes through the t-r cavity and reaches the crystal mixer even though the spark gap is fired. This leakage power has two components. One component is independent of the peak transmitter power; the other is directly proportional to, and of the order of, 66 db below the transmitter power (that is, about one eight-millionth of the transmitted power). In addition, most magnetrons generate appreciable power at multiples of the carrier frequency. The t-r cavity presents very little attenuation at these high frequencies. When the gap is fired, the t-r cavity may be considered as two small waveguides in parallel, providing paths from the input to the output aperture around the center electrodes. The dimensions of these guides are too small to transmit appreciable power at the proper frequency of the system but not small enough to cause appreciable attenuation of power at harmonic frequencies. The components of leakage power indicated above are not excessive provided the peak power of the transmitter is less than about 500 kw, and for such systems the t-r device shown in Fig. 50*A* gives adequate crystal protection. On the other hand, for peak power greater than about 500 kw, the leakage power may become so large that, for adequate crystal protection, the transmit-receive devices require additional refinement.

Figure 51*A* is an example of a transmit-receive device used in some high-power systems, and Fig. 51*B* shows its equivalent circuit. In addition to the t-r and anti-t-r devices, a pre-t-r tube is inserted at the T junction in the branch of the waveguide leading to the receiver. This pre-t-r tube is a quarter-wavelength section of waveguide. As shown in Fig. 51*A*, its ends are partially closed by thin metal plates, leaving rectangular

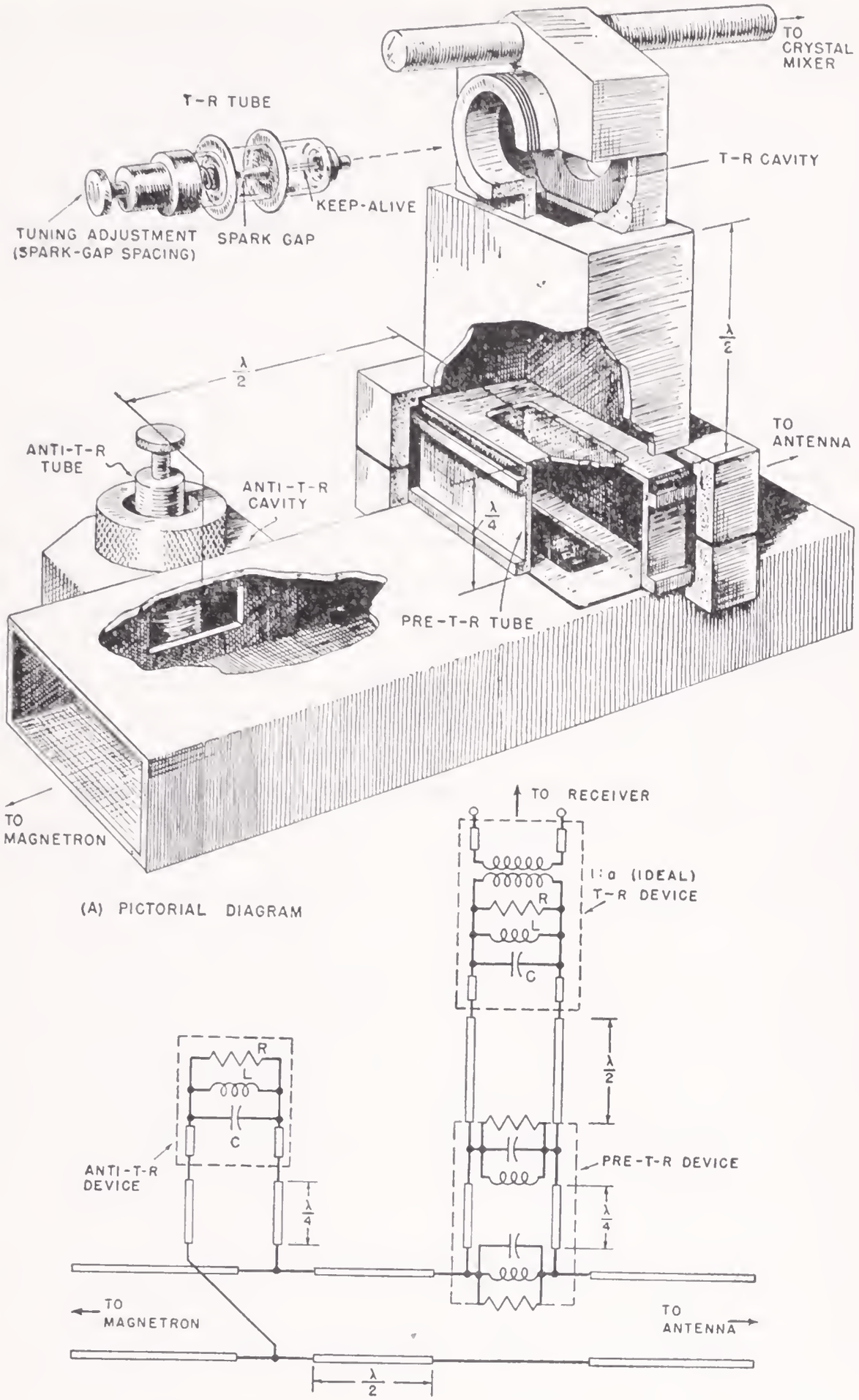


FIG. 51. Example of a t-r device used in high-power waveguide systems.

apertures somewhat smaller than the cross section of the guide. The aperture dimensions are proportioned so that the apertures are resonant at the transmitter frequency, and the openings are covered with low-loss glass to permit the space between the windows to be kept at a low pressure. The electric-field intensity within and near the resonant aperture is greater than that in the adjacent waveguide. This step-up in electric-field intensity is a measure of the resonant transformation of the aperture. The increased electric-field intensity, together with the reduced gas pressure, makes it possible to establish a glow discharge within the tube near one of the rectangular apertures.

During transmission, the electrical operation of the anti-t-r device is identical with that of Fig. 50*A*. The operation of the t-r device is, however, slightly different. Energy reaching the receiver T junction enters the branch of the waveguide feeding the t-r device and receiver. The t-r tube fires first, producing an essentially short circuit at the input aperture of the t-r device. This short circuit appears at the T junction, three quarter wavelengths away, as an open circuit. As a result of the voltage division between this open circuit and the relatively low impedance in the direction of the antenna (see the series connection in Fig. 51*B*), essentially all of the incident transmitter voltage appears, for a short time, across the input resonant aperture of the pre-t-r tube. This high voltage causes a glow discharge to occur across the input aperture, effectively short-circuiting the receiver branch of the waveguide and permitting power to flow to the antenna. The high impedance observed at the T junction in the direction of the receiver is in parallel with this short circuit, permitting only a small amount of carrier-frequency leakage power to flow toward the t-r tube and a still smaller component to reach the crystal mixer. The conducting "wall" at the resonant aperture offers appreciable attenuation to the higher frequency power as well, thus protecting the crystal mixer from nearly all components of leakage power. The slight delay in firing the pre-t-r tube is of no serious concern because the high-frequency content of the leakage power during the very early stages of the pulse is negligible. Only the t-r tube is provided with a keep-alive voltage. This is to ensure that the t-r tube will fire at the *earliest* possible moment during the transmitter pulse.

During the reception of echoes, the circuit behavior is similar to that discussed for other systems. The anti-t-r tube effectively short circuits the main transmission line (see Fig. 51*B*) when it is not fired and, consequently, is placed at an electrical distance of a half wavelength from the receiver branch. Because the apertures and the t-r tube are resonant when not fired, they do not interfere appreciably with the flow of echo power to the receiver. The t-r and anti-t-r cavities are tuned by changing the effective capacitance at the center of the cavity.

The output window of the pre-t-r tube is used only to maintain the low pressure in the tube. The quarter-wave spacing between windows is chosen in order that reflections at the two windows may cancel one another as nearly as possible during reception.

At 3 and 1 cm, the t-r cavity is made an integral part of the t-r tube. At these shorter wavelengths, the dielectric loss resulting from the presence of a glass envelope in the cavity would be so high that an appreciable loss of echo-signal strength would occur in the cavity. Assembling the gap and cavity as a unit eliminates the need for glass except as coverings for the coupling apertures.

A 3-cm t-r-tube and cavity assembly is illustrated in Fig. 52. The cavity is located in a copper block and is cylindrical in shape. The spark-gap electrodes are centered on the axis of the cylinder, and the keep-alive electrode extends from one spark-gap electrode to the top of an evacuated glass envelope mounted on the copper block. The glass envelope serves to increase the total volume of the evacuated space and thus aids in maintaining proper pressure within the cavity for long periods of time. Coupling to the cavity is provided by two circular, glass-covered apertures—one in each face of the block. The cavity is tuned by varying the spacing between spark-gap electrodes.

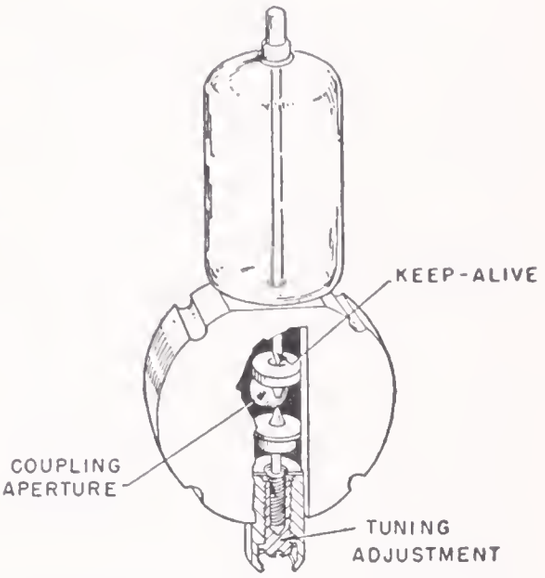


FIG. 52. A 3-cm t-r-tube and cavity assembly.

15. Input Circuits of Ultrahigh-frequency Receivers. Echo signals, after passing through the unfired t-r device in the r-f system, are applied to the input circuits of the receiver. A careful choice of input circuit is very important, because it is in this portion of the receiver that noise has the greatest effect on receiver sensitivity. In the microwave region, a low noise figure is obtained most readily by applying the echo signals directly to a mixer circuit employing a crystal, whereas in the uhf range it is usually advantageous to use a vacuum-tube preceded by one or two stages of radio-frequency amplification.

Vacuum tubes have several advantages over crystals in the input circuits of a radar receiver. Because crystals are delicate, they must be protected extremely well by the t-r device in order to avoid burnout during transmission. Vacuum tubes, on the other hand, are able to withstand relatively large amounts of transmitter power. Mixer circuits employing triodes also have a somewhat higher conversion gain than crystal mixers, although they generate more noise. The effect of the

increased noise may be offset at ultrahigh frequencies by preceding the triode mixer by one or more stages of radio-frequency amplification—the noise figure of a triode is appreciably less when used as an amplifier than when used as a mixer (see Art. 29, Chap. VI).

Unfortunately, the tubes that are available for operation in the microwave region generate an excessive amount of noise and are therefore not suitable for the input circuits of microwave receivers. Consequently, use of vacuum-tube input circuits is restricted to receivers that operate in the uhf range. Vacuum-tube input circuits are discussed in this article, and several forms of crystal-mixer input circuits are presented in the next two articles.

The input circuits of uhf receivers usually employ disk-seal triodes (Art. 1, Chap. X) and are frequently connected to form a grid-return type of circuit. An r-f amplifier employing a disk-seal tube in a grid-return circuit is very stable and has very little tendency to oscillate, the only feedback path being through the extremely low plate-to-cathode capacitance of the tube. Pentodes are available for operation in the uhf range, but they generally have a higher noise figure than triodes and are therefore less suitable in radar applications.

In uhf receivers that operate at 200 Mcps or below, the tuned circuits of the input stages are composed of lumped inductances and capacitances. Often physical capacitors are unnecessary, the inductors being made to resonate with the tube and wiring capacitances. In the region of 1,000 Mcps the lumped-parameter resonant elements are replaced by cavity resonators.

An example of an input circuit of a receiver that operates in the vicinity of 1,000 Mcps is given in Fig. 53*A*. The circuit consists of a single stage of radio-frequency amplification, a local oscillator, and a mixer. Lighthouse tubes and coaxial-line resonators are employed in each unit, and connections between units are made by means of short sections of coaxial cable.

The r-f amplifier is similar in construction to the lighthouse-tube oscillator of Fig. 16, Chap. X, the only important difference being that in the amplifier the middle cylinder of the tuner is not cut away at the grid terminal of the tube. The extra coupling needed in the oscillator between plate and cathode lines is thus eliminated, and the amplifier plate and cathode circuits are isolated completely except for the small interelectrode capacitance between plate and cathode of the tube. The plate supply voltage is conducted to the plate terminal of the amplifier tube V_1 by means of an insulated wire which passes through the innermost coaxial cylinder. This cylinder, which serves as one conductor of the plate line, is coupled to the plate terminal through capacitor C_1 in order that the tuner may be operated at ground potential. Inductor L_1 serves

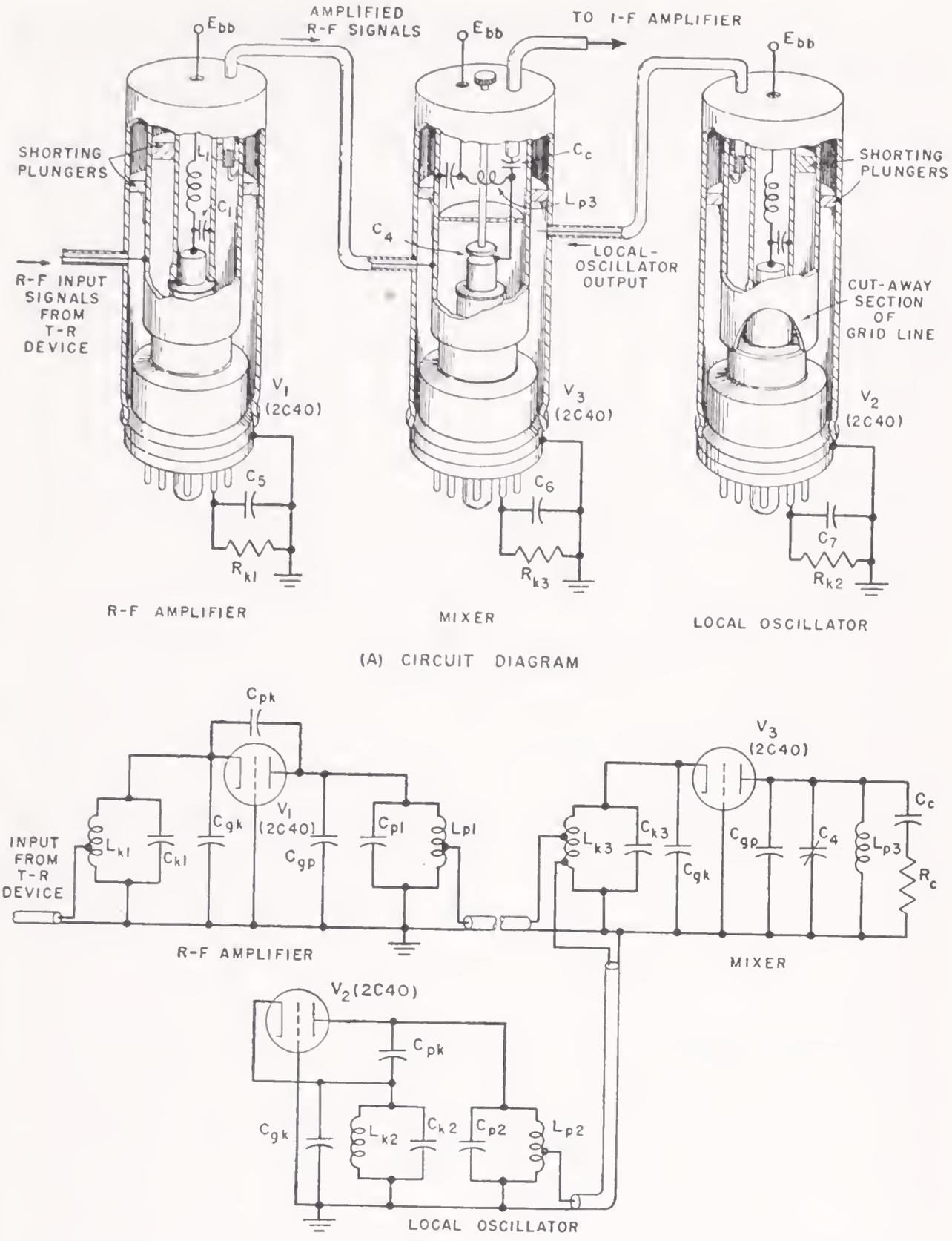


FIG. 53. Example of an input circuit of a uhf receiver.

as a choke to prevent r-f signals from entering the power supply. The middle cylinder of the tuner connects directly to the grid terminal, and the outer cylinder fits over the shell of the tube, which is grounded. All cylinders are connected together by the shorting plungers in the plate and cathode lines. A by-pass capacitor within the tube provides an r-f connection between shell and cathode, and cathode bias voltage is developed

across C_{k1} and R_{k1} in parallel between the cathode pin and ground. Input signals from the t-r device are coupled into the cathode-to-grid line by means of a probe, and a coupling loop is used to take the amplified r-f signals from the plate-to-grid line.

An r-f equivalent circuit of the amplifier, in which the plate and cathode lines are replaced by parallel L-C circuits, is shown in the upper left portion of Fig. 53B. Observe that the equivalent circuit is drawn so as to emphasize the grid-return aspect of the amplifier. The plate line, represented by L_{p1} and C_{p1} in Fig. 53B, is adjusted to resonate with the plate-to-grid capacitance C_{gp} of the tube at the frequency of the incoming signal, and the cathode line in parallel with C_{gk} is adjusted very nearly to the frequency of the incoming signals. Actually, the cathode line is made slightly inductive in order that any feedback voltage supplied through the small plate-to-cathode capacitance C_{pk} (approximately 0.05 $\mu\mu$f) will have a phase opposite to that required to start oscillations. Thus the cathode-line length is somewhat less than a quarter wave at the signal frequency.

Because the output coupling loop of the amplifier is located near the shorting plunger in the plate line, the output coaxial cable is shown as a tapped connection to L_{p1} in Fig. 53B. The plate line thus acts as a resonant transformer to change the low input resistance of the cable (approximately 50 ohms, since the cable is usually matched at the far end) to a higher resistance across the entire resonant circuit. The value of load resistance across the resonant circuit is governed primarily by the bandwidth required for the amplifier. Since no provision is made for tuning the amplifier during its normal operation, the bandwidth must be wide enough to pass echo signals despite slight changes in the transmitter frequency.

The tuned cathode circuit is loaded by the input resistance of the amplifier in parallel with a resistance representing grid-loading effects and tuned-circuit losses. The net resistance is matched approximately to the characteristic resistance of the input line by tapping the input cable to a point near the shorted end of the cathode line. The tap location is chosen for optimum signal-to-noise ratio and results in a resistance termination for the input line which is slightly more than its characteristic resistance (see Art. 17, Chap. VI). Calculations of input impedance and gain of the r-f amplifier may be carried out in a manner analogous to that used to determine similar quantities for the grid-return stage of the i-f amplifier discussed in Art. 25 of Chap. VI.

The local oscillator, shown at the right in Fig. 53A, is similar in construction to the r-f amplifier except that the grid cylinder is partly cut away near the local-oscillator tube. Thus C_{pk} in the oscillator equivalent circuit (see Fig. 53B) includes both the plate-to-cathode capacitance of

the tube and the external coupling capacitance resulting from the opening in the grid line. All cylinders of the oscillator tuner, like those of the r-f amplifier, operate at ground voltage, and the cathode is maintained at a positive potential with respect to ground because of the bias voltage developed across C_{k2} in parallel with R_{k2}.

The local oscillator is tuned in accordance with the principles discussed for the lighthouse-tube oscillator in Art. 7 of Chap. X. The plate-line length is the primary control for frequency and, in fact, constitutes the main tuning control for the receiver. Tuning is accomplished by means of a knob or dial connected to the plate-line shorting plunger through a flexible cable (not shown in the figure). The cathode-line length governs the magnitude and phase of the voltage fed back to the cathode circuit from the plate circuit and must be adjusted so that a capacitive reactance appears between grid and cathode at the oscillation frequency. At frequencies in the neighborhood of 1,000 Mcps, it is mechanically feasible to choose line lengths so that each line employs its lowest mode of resonance. Thus the cathode line is approximately one-quarter wave at the oscillator frequency, and the plate line is less than a quarter wave.

Power coupled from the plate line of the local oscillator is applied, together with echo signals from the r-f amplifier, to the cathode line of the mixer tube V_3. A component of current at a frequency equal to the difference between the local-oscillator and echo-signal frequencies is produced in the plate circuit of the mixer in accordance with the principles discussed in Art. 29 of Chap. VI, and the plate circuit is tuned to this difference frequency.

The mixer tuner differs from the local-oscillator and r-f amplifier tuners in that it employs a coaxial-line cavity only between cathode and grid, lumped parameters being used in the plate circuit. The tuned plate circuit consists of inductor L_{p3} in parallel with interelectrode capacitance C_{gp}, the series combination of C_c and the input resistance R_c of the output line, and the variable capacitance C_4 formed by a movable metallic disk near the plate terminal of the tube. The capacitance of C_4 is varied by changing the disk position by means of a rod which passes through the end of the tuner.

The various components comprising the tuned plate load of the mixer are illustrated in Fig. 53*B*. Because input signals are applied between grid and cathode of the mixer and the intermediate-frequency signals appear between grid and plate, the equivalent circuit is represented as a grid-return circuit. From the equivalent circuit it may be seen that the variable capacitor C_4 serves to tune the remaining plate-circuit elements to resonance at the intermediate frequency. Placing capacitor C_c in series with the output cable isolates the cable from the power supply voltage and causes the low input impedance R_c of the cable to appear as a

higher resistance across the entire circuit. (The impedance relations for such a connection are similar to those discussed for the output connection illustrated in Fig. 44*B*, Chap. VI.)

The receiver input circuit of Fig. 53*A* requires that the interconnecting cable between r-f amplifier and mixer be short. Otherwise appreciable attenuation of echo signals will occur in the cable and cause noise generated in the mixer to become important. If cable loss is excessive, it may become necessary to employ a second stage of r-f amplification in order to maintain the signal level well above the noise level of the mixer.

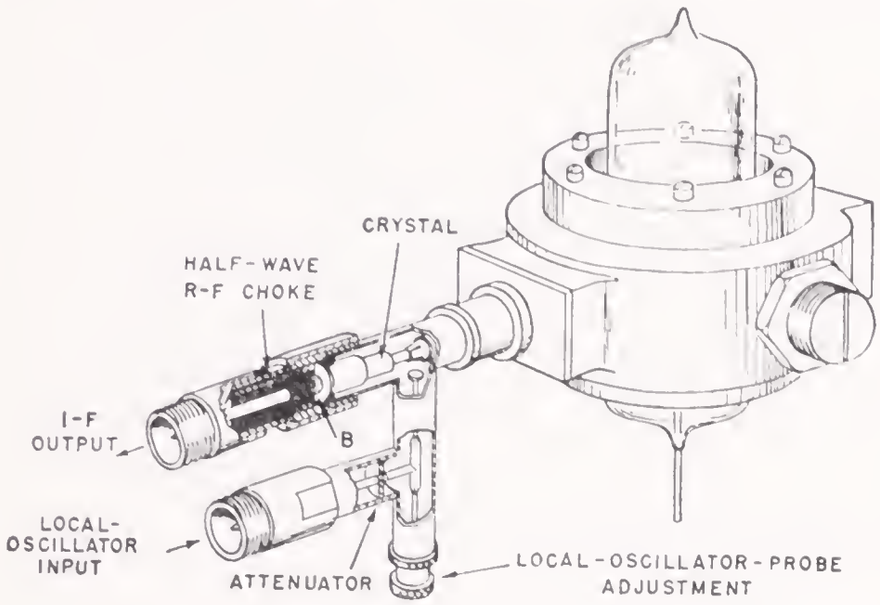
16. Single-ended Crystal Mixers. In microwave receivers, amplification of echo signals at the carrier-frequency level is in general undesirable because of the rather large amount of noise generated by available microwave-amplifier tubes. Echo signals, after passing through the t-r device, are therefore applied directly to the crystal mixer. A voltage from a local oscillator, usually a klystron, is applied to the mixer simultaneously with the echo signals, and signals at the intermediate frequency are derived at the mixer output terminals.

The important features of a 10-cm mixer employing coaxial-line sections are given in Fig. 54*A*. Echo signals are coupled into the transmission-line section between t-r cavity and crystal by means of a coupling loop which penetrates a short distance into the cavity. The crystal fits snugly into the inner conductor of the line and is held concentric with the outer conductor by means of a dielectric bead. The local-oscillator signal is introduced into the mixer assembly by means of a coupling probe located approximately a quarter wave (at the local-oscillator frequency) from the coupling loop in the t-r cavity. A high impedance in the direction of the cavity is thus presented to the local-oscillator signal at the probe. The strength of the local-oscillator signal at the crystal may be changed by varying the distance the probe extends into the main line. Power from the local oscillator is conducted to the probe over a section of transmission line.

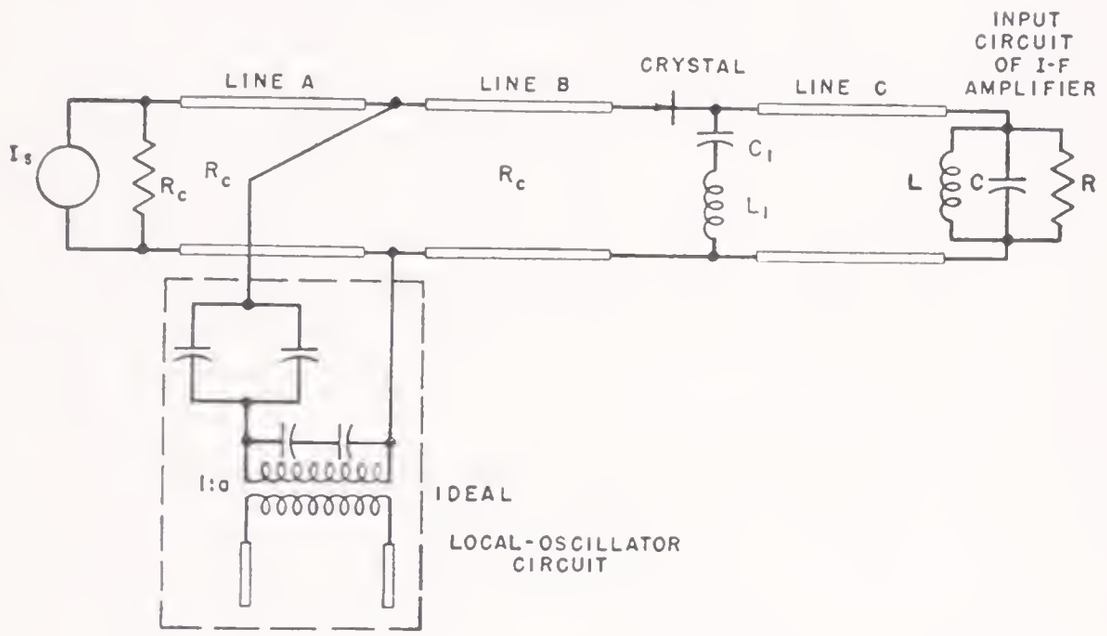
Connection between the crystal and inner conductor of the line section carrying i-f signals is provided by a spring contact. The half-wave r-f choke appearing on the i-f side of the crystal provides an electrical short circuit for radio-frequency signals at point *B*. The d-c return path required for crystal current is provided at the r-f end of the mixer assembly by the t-r coupling loop and at the i-f end by the shunt-inductance element located in the grid circuit of the input i-f-amplifier stage.

A two-wire equivalent circuit of the mixer is given in Fig. 54*B*. If small losses in the r-f line and t-r cavity are ignored, the portion of the r-f transmission system between antenna and output terminals of the t-r cavity may be represented by a signal-current source I_s in shunt with the characteristic resistance R_c of the transmission line. Because the local

oscillator is loosely coupled to the main line, negligible admittance appears in shunt with the lines A and B in Fig. 54B. Elements L_1 and C_1 represent the r-f choke at the i-f end of the crystal and are series resonant at the echo-signal frequency. Hence line B is terminated by the crystal at the echo-signal frequency. For maximum receiver sensitivity



(A) PICTORIAL REPRESENTATION



(B) AN EQUIVALENT CIRCUIT

FIG. 54. A 10-cm mixer.

it is important that essentially all the echo power available at the t-r device be transferred to the crystal. The crystal should therefore present a matched-load termination to the main-line section. Fortunately, 10-cm crystal units, when operated with $\frac{5}{16}$-in. 50-ohm coaxial lines, provide essentially a matched r-f load for the line. An impedance-changing device between crystal and t-r cavity is therefore unnecessary.

An equivalent signal circuit for the mixer may be obtained from Fig. 54*B* by replacing the crystal by its equivalent linear network (see Fig. 53 of Chap. VI). The signal circuit may then be used to relate the incoming r-f signals to the intermediate-frequency output signals in accordance with the principles discussed in Art. 28 of Chap. VI.

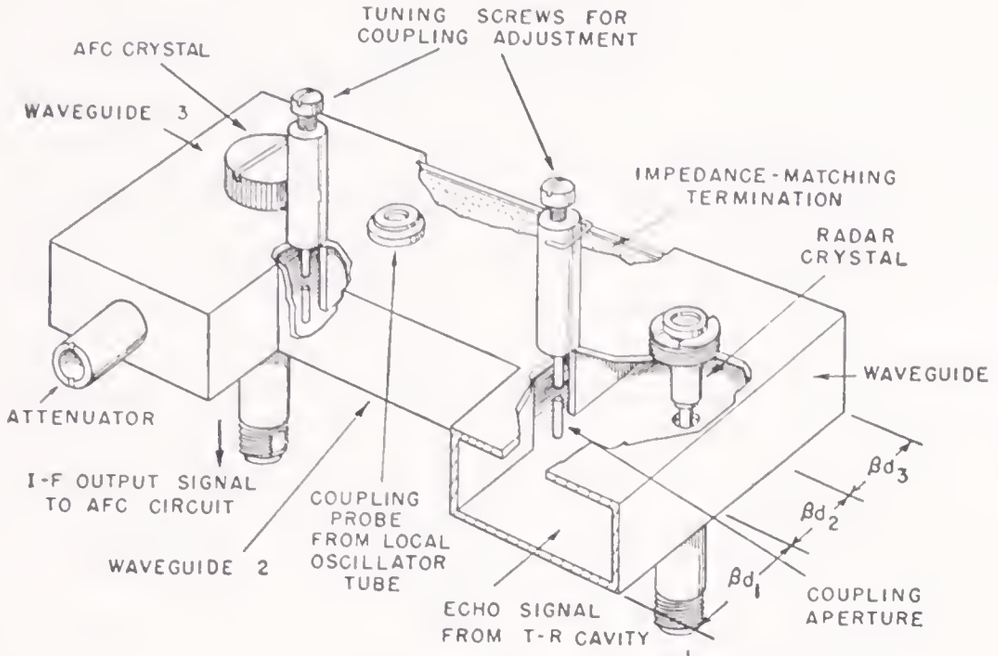
Because the local oscillator is loosely coupled to the main line containing the crystal, an appreciable mismatch of impedances exists on the local-oscillator line at the coupling probe. A large impedance mismatch is in general undesirable because of the wide variations in local-oscillator load impedance (and hence power output) that result when the oscillator frequency is varied. A good impedance match is particularly important if the line between local oscillator and probe is long. To ensure a relatively constant load impedance for the oscillator, a resistance-disk type of attenuator (see Art. 8) is oftentimes inserted in the local-oscillator line, as shown in Fig. 54*A*, in order to terminate the line section between oscillator and disk in its characteristic resistance. An alternative to the resistance disk is a piece of lossy cable which may be inserted between local oscillator and mixer assembly.

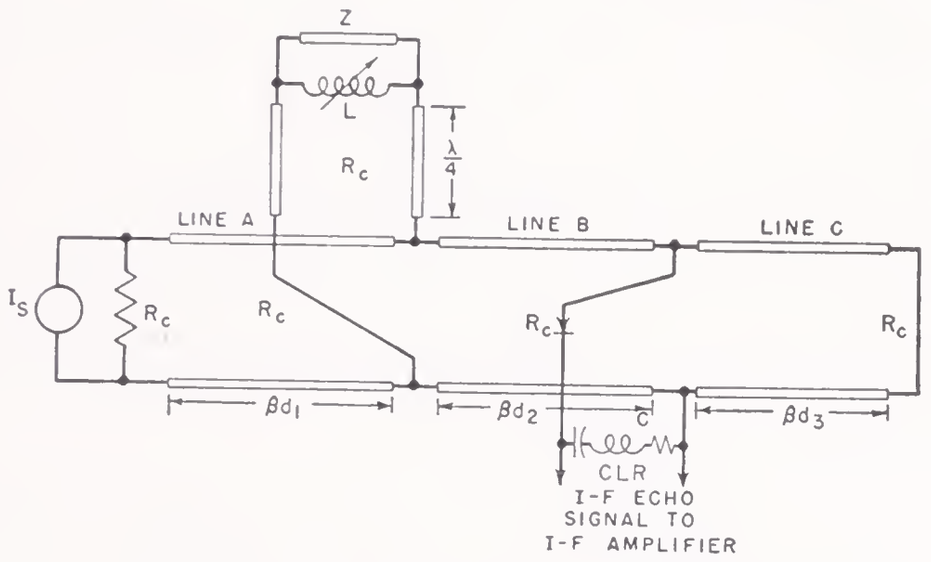
Oftentimes mixer assemblies include, in addition to the circuit that supplies i-f echo signals, a separate mixer for generating i-f signals for the automatic-frequency-control circuit. An example of a 3-cm mixer assembly of this type is shown in Fig. 55*A*. Local-oscillator power and r-f echo signals from the t-r cavity are coupled into waveguide 1, which serves as the echo-signal channel. A crystal extends between top and bottom walls of the guide, the lower end of the crystal being isolated from the guide by means of a dielectric washer. Echo-signal voltages in the intermediate-frequency band appear between the lower end of the crystal and the waveguide and are conducted to the input circuit of the i-f amplifier, usually over a coaxial cable.

Waveguide 3 is similar in construction to waveguide 1 and serves as the afc channel. Power from the local oscillator and a small amount of the transmitter power are applied to the afc crystal, and i-f signals for the afc circuit are derived at the lower end of the crystal mount, as before. Attenuation between the main guide and guide 3 is provided by means of a circular-waveguide section operating below cutoff. Sometimes a disk type of attenuator is included in the circular section to reduce further the amount of transmitter power reaching the mixer assembly.

The local-oscillator tube is mounted above waveguide 2 in a manner that permits the local-oscillator output probe to extend a short distance into guide 2. The coupling probe may be displaced slightly from the center of the transverse axis of the guide in order to obtain a load impedance for the oscillator which yields a reasonably uniform power output over the tuning range.

Local-oscillator power is coupled from guide 2 into guides 1 and 3 through coupling apertures located in the side walls of the guides. A matching post and tuning screw extending from top and bottom walls of the guides at the aperture locations serve to regulate the amount of local-oscillator power coupled to the mixer circuits.





(B) AN R-F EQUIVALENT CIRCUIT OF ECHO-SIGNAL CHANNEL

FIG. 55. A 3-cm mixer.

For optimum performance of the radar system, the echo-signal channel of the mixer should provide a matched-load impedance to the waveguide transmission system. The elements that determine the mixer input impedance are shown in the equivalent circuit of Fig. 55*B*. The equivalent circuit is derived by dividing guide 1 into three sections of lengths βd_1, βd_2, and βd_3 and representing the sections by equivalent lengths of transmission lines (lines *A*, *B*, and *C* in Fig. 55*B*). The signal-current

generator I_s in shunt with R_c in the equivalent circuit accounts for the portion of the r-f system between antenna and mixer. The coupling aperture between local-oscillator and mixer guides (guides 1 and 2) is represented by a variable inductance L, and in shunt with L is an impedance Z which accounts for the impedance seen in the direction of the local-oscillator guide. The parallel combination of L and Z, in accordance with Art. 18 of Chap. IX, terminates a quarter-wave section of line which is in shunt with lines A and B. At the echo-signal frequency, the impedance of L and Z in parallel is maintained small enough in comparison with the characteristic resistance of the quarter-wave line to make the input impedance of the quarter-wave section extremely high and hence the amount of echo-signal power entering the local-oscillator guide extremely small.

The crystal is located at the junction of lines B and C, as shown in Fig. 55B, and across the crystal output terminals is a series R-L-C circuit. The counterpart of the R-L-C combination in Fig. 55A is an r-f choke in the i-f output lead of the mixer. The choke prevents r-f signals from entering the input circuit of the i-f amplifier.

The length of shorted-end line C must be such that the input admittance of line C, in shunt with the r-f admittance of the crystal, provides essentially a matched load for line B. Varying the length of line C changes the susceptance component of the total admittance at the crystal location. A length of approximately $\lambda/4$ is usually preferable in order to minimize changes in admittance with frequency.

In some mixers, with line C approximately a quarter wave long, it is possible to cancel susceptive components of the admittance at the crystal, but because of incorrect conductance components, an impedance match is not possible. In this event, the crystal may be shifted to a point nearer one side wall of the guide. Moving the crystal along the transverse axis of the guide has the effect mainly of varying the conductance component of the admittance in the plane of the crystal.

When the radar system is inoperative, it is possible that an appreciable amount of r-f power from near-by radar sets may reach the crystal mixer and thus cause crystal burnout. It is difficult to fire the t-r gap under these conditions because of the absence of keep-alive voltage. In order to prevent crystal damage from this cause, mixers often include some sort of auxiliary protective device which is effective whenever the set is not being used. In waveguide systems, the protective device may take the form of a solenoid-operated metallic flap or *shutter*. The shutter enters the guide from a side wall and slides completely across the guide whenever the set is turned off. When power is applied to the system, a solenoid serves to withdraw the shutter, thus permitting echo power to reach the crystal. A similar arrangement is used in coaxial-line systems, except

that the shutter enters the t-r cavity in a way that isolates the output coupling loop from the rest of the cavity. In coaxial-line systems, it is also possible to shunt the main-line section between t-r cavity and crystal with a half-wave shorted-end stub equipped with a solenoid-operated short-circuiting strip at the mid-point of the stub. When the set is inoperative, the solenoid is unenergized and the shorting strip is removed from the stub. The stub is therefore a half wave long, and in effect a short circuit is placed across the main line at the stub location. Energizing the solenoid causes the shorting strip to enter the stub at the mid-point, thus making the stub effectively a quarter wave long. The quarter-wave stub has negligible effect on echo signals passing to the receiver.

17. Balanced Crystal Mixers. The noise introduced into the input circuit of a microwave receiver by the local oscillator may be reduced appreciably through use of a balanced mixer. An input circuit of this type employs two crystal mixers connected so that the local-oscillator voltages applied to each crystal are in phase and the echo-signal voltages at each crystal are 180 deg apart in phase. As explained below, this method of circuit connection causes the i-f components of local-oscillator noise to cancel in the i-f-amplifier input circuit and the i-f components of signal to add. Because of their ability to suppress local-oscillator noise and because of certain additional advantages to be discussed later in this article, balanced mixers are oftentimes preferable to single-ended mixers.

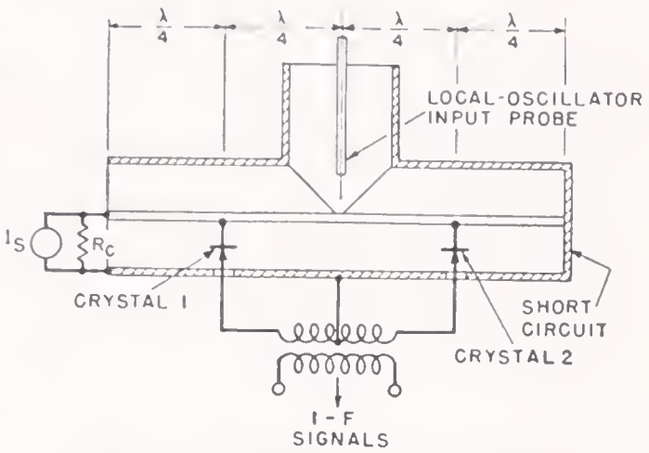


FIG. 56. Balanced-mixer circuit.

The basic principles of a balanced mixer may be explained from the diagram of Fig. 56, which shows two crystals mounted a half wavelength apart (at the signal frequency) between conductors of a coaxial-line section. Each crystal is located a quarter wave from the nearer end of the line, and the local-oscillator signal is probe-coupled to the line at a point midway between the crystals. The output terminals of the crystals are connected to an i-f transformer, the primary winding of which is balanced to ground. The signal-current generator I_s in shunt with resistance R_c accounts for signals entering the mixer from the t-r device.

The local-oscillator coupling probe gives rise to two waves traveling in opposite directions in the coaxial line Assume that at the local-oscillator frequency the t-r device presents a short circuit to the local-oscillator wave traveling to the left in Fig. 56 and that the crystals have identical r-f impedances equal to twice the characteristic resistance of the line.

Then equal-amplitude, in-phase, local-oscillator voltages appear at the crystals. Because of the 180-deg spacing between crystals, the r-f echo-signal voltages at the crystals are also of equal amplitude but are 180 deg apart in phase. From Fig. 52 of Chap. VI, it may be seen that an in-phase relation between local-oscillator and echo-signal voltages produces an i-f signal voltage of one phase and that a 180-deg phase relation between local-oscillator and echo-signal voltages produces an i-f signal voltage of opposite phase. Thus the two i-f echo-signal voltages add in the balanced input circuit of the i-f amplifier. Because the noise that accompanies the local-oscillator signal produces voltage components that bear the same phase relation to the local-oscillator voltage at each crystal, the corresponding i-f components of noise at the output of each crystal are in phase and hence produce no resultant noise voltage across the i-f transformer.

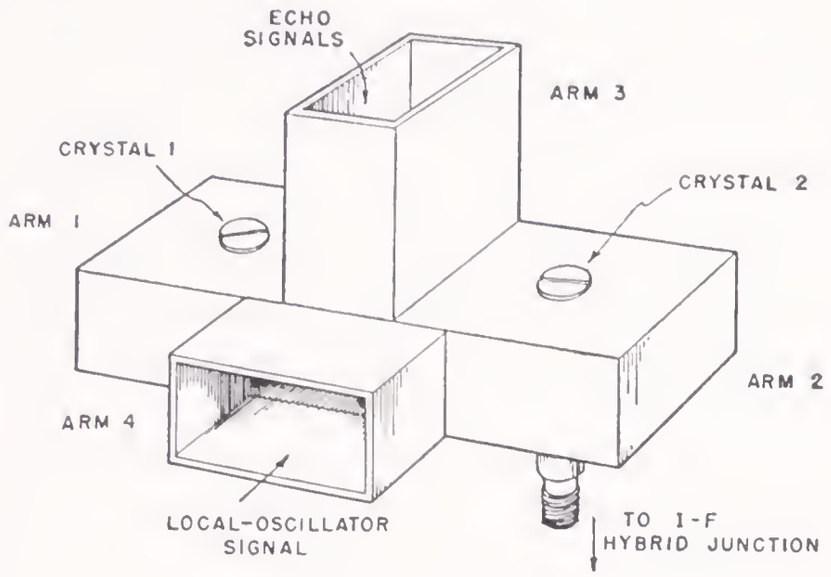
In order to realize full benefit from use of a balanced-mixer circuit, the crystals should have similar r-f impedances as assumed above, and in addition they should have similar conversion gains. Under these conditions of operation, only one-half the available r-f power is applied to each crystal, but since the i-f signal powers from the crystals add together, the total i-f power is the same as that obtained from a single-ended mixer. Since i-f components of local-oscillator noise are completely eliminated, maximum reduction in over-all noise figure is obtained.

An unbalance in conversion gain yields an unbalance in the i-f signal voltages obtained from each crystal and also an unbalance in the i-f components of local-oscillator noise voltages at the crystal output terminals. However, since the i-f signal voltages are additive in the crystal output circuit and the local-oscillator-noise components are subtractive, small unbalances in conversion gains do not seriously affect the over-all receiver noise figure.

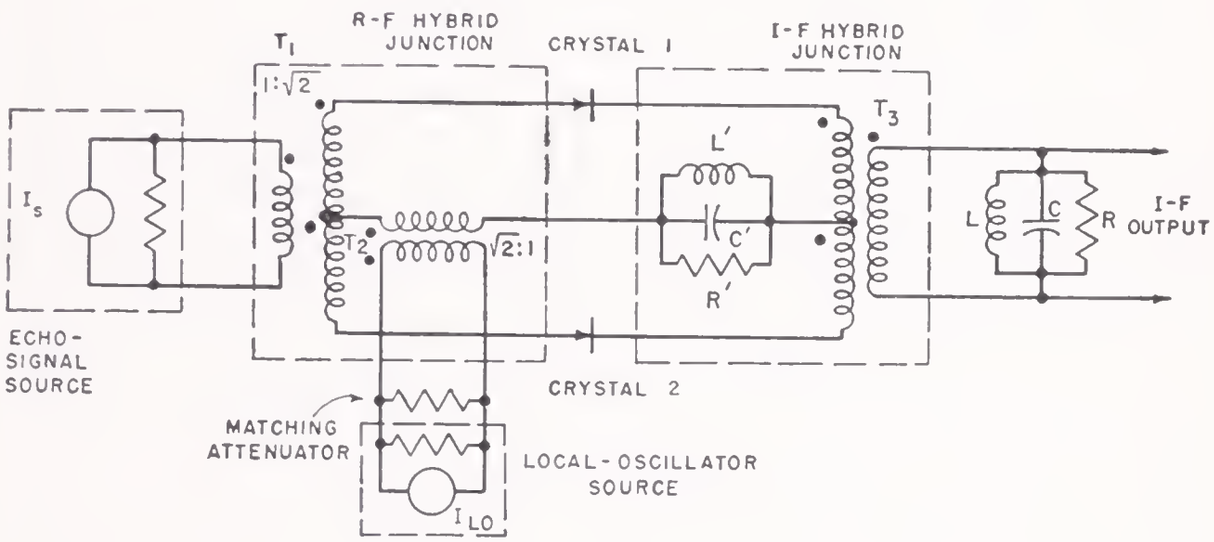
More important are the effects of a difference in the r-f admittances of the crystals. An unbalance of this type causes both the echo-signal power and the local-oscillator power to divide unequally between the two crystals. As a result, the i-f signal power obtained from the mixer is appreciably reduced, and in addition the i-f components of local-oscillator noise are prevented from canceling completely in the i-f input circuit. Crystals intended for use in a mixer constructed as shown in Fig. 56 must be carefully selected on the basis of equal r-f admittances in order to ensure most satisfactory operation.

A balanced mixer that is superior in several respects to the coaxial-line mixer just described is shown in Fig. 57*A*. This mixer employs a waveguide hybrid junction of the type discussed in Art. 6. Crystals are located in one pair of entries (arms 1 and 2), and echo signals and local-oscillator signals are applied at the remaining entries (arms 3 and 4).

Usually in a mixer of this type the input circuit of the i-f amplifier is also arranged in the form of a hybrid junction. Details of connections of the i-f amplifier input circuit, together with the ideal-transformer equivalent circuit of the waveguide junction (see Fig. 17), are given in Fig. 57*B*. The parallel circuit comprising L, C, and R is tuned to the intermediate



(A) PICTORIAL DIAGRAM



(B) EQUIVALENT CIRCUIT

FIG. 57. Balanced mixer employing hybrid junctions.

frequency, and the impedance of the parallel combination of L', C', and R' is made equal to the conjugate of the i-f impedance appearing between the terminals to which L', C', and R' are connected.

Ideally the two crystals should be alike. Furthermore, all r-f impedances should terminate the entries of the r-f hybrid junction in matched loads, and all i-f impedances should similarly terminate the entries of the i-f hybrid junction. Then, because of the properties of a matched hybrid junction local-oscillator signals at the crystals are equal in amplitude and

in phase, and echo signals at the crystals are equal in amplitude and 180 deg apart in phase.

The i-f components of signals from the crystals are applied in push-pull to ideal transformer T_3 and appear across L, C, and R but not across L', C', and R'. The i-f components of noise, on the other hand, are in effect applied between end terminals of T_3 in parallel and appear across L', C', and R' but not across the i-f output terminals. Local-oscillator noise is thus suppressed completely. Because the r-f hybrid junction is matched, the total available power from the echo-signal and local-oscillator arms divides equally between the two crystals. As in the coaxial-line mixer of Fig. 56, the i-f power from each crystal is one-half the available echo-signal power times the conversion gain G of the crystal. Since i-f power from each crystal is delivered at the i-f output terminal, the total i-f output is the product of the echo-signal power and G, as before.

Ordinarily, in single-ended mixers and in the coaxial-line balanced mixer of Fig. 56, the local oscillator must be loosely coupled to the mixer in order to prevent an appreciable amount of echo-signal power from entering the local-oscillator circuit. As a result, the local oscillator must generate much more power than is actually needed for mixer operation, the excess power being dissipated in an impedance-changing device located in the local-oscillator circuit. Because in a hybrid-junction mixer the echo-signal and local-oscillator arms are decoupled, there is no tendency for echo signals to enter the local-oscillator circuit, and hence there is no need for loose coupling between local oscillator and mixer. A local oscillator with relatively low power output can therefore be used, the only power necessary being that which is needed to drive the crystals. Actually most local oscillators are capable of generating more than the minimum power required, so that a dissipative attenuator can be inserted in the local oscillator in order to ensure a matched termination for the arm. Such an attenuator is desirable, as will be seen later.

Decoupling between echo-signal and local-oscillator arms also means that there is no tendency for local oscillator power to enter the echo-signal arm. Rejection of local-oscillator signal by the echo-signal arm is particularly advantageous in systems employing a broad-band t-r device which would allow local-oscillator power to pass through and be radiated by the antenna.

Another advantage of the hybrid-junction mixer is that its operation is less sensitive to differences in the r-f impedance of the crystals than is the balanced coaxial-line mixer. Because the two crystal arms are decoupled, any r-f power reflected from one crystal cannot reach the other crystal. All the reflected power is directed into the echo-signal and local-oscillator arms and absorbed by the matched terminations. Thus the amount of

r-f power delivered to one crystal is independent of the r-f impedance of the other.

If a mismatch should exist in either the echo-signal or local-oscillator arm, a reflection of an r-f signal arriving at the mismatched termination will occur, and the signal reflected from the termination will enter the two crystal arms. Under these circumstances, the two crystal arms are not completely decoupled. At the echo-signal frequency, a mismatch in the echo-signal and local-oscillator arms is unlikely because in general the t-r device provides a matched load for the echo-signal arm, and as indicated above, a matching attenuator can normally be inserted in the local-oscillator arm.

In general if the crystals provide an impedance match at the echo-signal frequency, they do not yield a perfect match at the local-oscillator frequency. This mismatch is unimportant, since the reflected power will be absorbed in the echo-signal and local-oscillator arms. The attenuator in the latter arm serves to prevent the returned signal from reacting on the oscillator. Should the echo-signal arm not provide a matched termination for the returned local-oscillator signal, further reflection of the signal will occur and the balance of local-oscillator power at the crystals will be further upset. The effect of such an unbalance is to change the conversion gains of the crystals slightly, but usually the change is negligible even with moderate amounts of mismatch at the local-oscillator frequency.

Use of the hybrid-junction type of i-f input circuit shown in Fig. 57*B* ensures suppression of local-oscillator noise despite unbalance in the i-f impedances of the crystals. When all entries of the i-f hybrid junction are terminated in a load that is the conjugate of the impedance the load faces, the two crystals, constituting one pair of entries, are decoupled and the two resonant circuits, constituting the other pair, are also decoupled. If the resonant-load entries remain matched, the amount of i-f signal power coupled from one crystal to the output terminals is independent even if the i-f impedance of the other crystal does not match the arms. Furthermore, the i-f components of noise cancel in the output circuit independently of the i-f crystal impedance as long as the crystals develop equal i-f noise voltage. The i-f components of noise appear only across the L', C', and R' arm. If crystals are tested for equal conversion gains while loaded by the conjugate of the i-f impedance of a typical crystal, equal i-f noise voltages can be expected when the crystals are employed in the circuit of Fig. 57.

Terminating the i-f output entry in a matched load does not in general ensure minimum noise figure. Mismatching the output entry in order to minimize the noise figure alters the load for the push-pull i-f echo signals and thus reduces the echo signal power delivered to the load but does not

alter the load for the i-f components of noise which are applied to T_3 in parallel. The local-oscillator noise-suppression property of the mixer is therefore retained.

18. Test Equipment. In this article several pieces of test equipment commonly used for making measurements on the r-f components of radar systems are described.

Standing-wave Indicators. As explained in Art. 10, the r-f transmission system of a radar set should operate at approximately unity standing-wave ratio during the transmission of pulses. Oftentimes it is desirable to determine this quantity experimentally while the set is in operation. In one method of making the measurement, two directional couplers are connected so that one coupler provides an indication of the power being propagated in the incident wave in the main line and so that the other gives an indication of the power in the reflected wave. If P^+ represents the incident power and P^- the reflected power in a lossless line,

$$\frac{P^-}{P^+} = \frac{|E^-|^2/R_c}{|E^+|^2/R_c} = |\Gamma|^2$$

where $|E^+|$ and $|E^-|$ are the magnitudes of the incident and reflected voltages, R_c is the characteristic resistance of the line, and $|\Gamma|$ is the magnitude of the reflection coefficient. Since the standing-wave ratio ρ is equal to $(1 + |\Gamma|)/(1 - |\Gamma|)$,

$$\rho = \frac{1 + \sqrt{\frac{P^-}{P^+}}}{1 - \sqrt{\frac{P^-}{P^+}}} \tag{5}$$

Thus if P^- and P^+ are measured, the corresponding standing-wave ratio may be calculated from Eq. 5. Since the concept of reflection coefficient may be applied to both two-wire lines and waveguides propagating a single wave type, Eq. 5 may be applied to either type of transmission system.

Figure 58 shows two two-hole, directional couplers arranged to measure the power in the incident and reflected waves in a rectangular waveguide transmission system. Coupler A responds to incident-wave power and coupler B to reflected-wave power. Signals are taken from each coupler by means of a pickup probe which extends from the end of the inner conductor of the output coaxial line into the coupler. Connected to the output line of each coupler is a detection device which serves to provide an indication of the magnitude of the incident or reflected power in the main guide.

A convenient form of detection device consists of a crystal detector in series with a microammeter. The crystal rectifies the r-f signal at the

crystal location, and the microammeter indicates the rectified current. The power in the main-line wave is determined from knowledge of the rectification characteristic of the crystal (the r-f power corresponding to

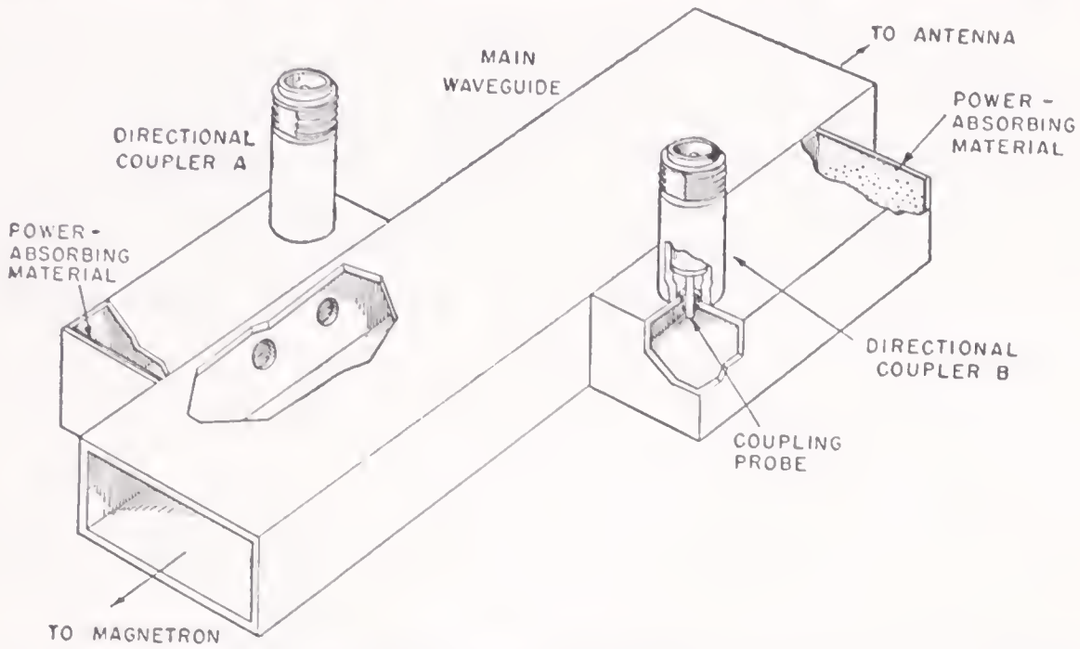


FIG. 58. Two two-hole, directional couplers arranged to provide an indication of the power in the incident and reflected waves in a waveguide.

a given value of rectified current) and the amount of attenuation between main line and crystal.

Figure 59 shows details of a crystal mount for use with the directional couplers of Fig. 58. The mount is essentially a short section of coaxial line, the inner conductor of which is the crystal. The output end of the crystal is by-passed for r-f signals to the outer conductor, whereas the rectified crystal current is directed to an output microammeter. If the crystal mount is used in conjunction with a probe type of pickup device, as in Fig. 58, a coaxial-line section containing a quarter-wave stub must be placed between the pickup probe and crystal mount in order to provide a return path for the rectified crystal current.

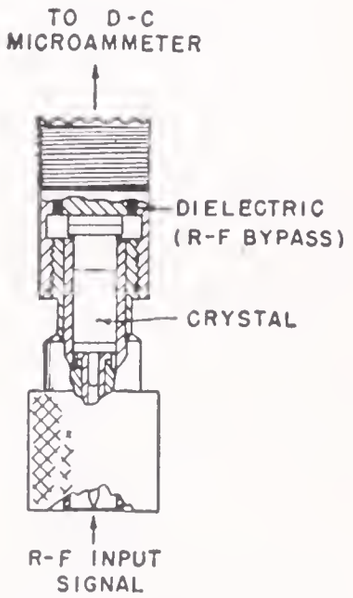


FIG. 59. Crystal mount for use with directional couplers of Fig. 58.

A device for measuring standing-wave ratio in coaxial lines is shown in Fig. 60*A*. The equipment consists of a slotted section of line and a detection device constructed so that it may be moved along the slot. The detection device indicates the magnitude of the voltage between conductors of the main line. Thus the maximum and minimum voltage between conductors may be determined and the standing-wave ratio calculated from the relation $\rho = |E|_{\max}/|E|_{\min}$. In order that both maximum- and minimum-voltage points may be located, the slot must be

at least a half wave long; usually its length is a full wavelength or more. Because the slot is cut parallel to the direction of current flow, the amount of power lost through radiation from the slot is kept small (see Art. 18, Chap. IX).

The detection device shown in Fig. 60*A* consists of a pickup probe that penetrates a short distance radially into the space between conductors of the slotted section and a section of coaxial line in which is mounted a detector element (see Fig. 60*B*). A voltage equal to the product of the electric-field intensity at the probe and the effective length of the probe is applied to the attached line. Because the electric field between main-line

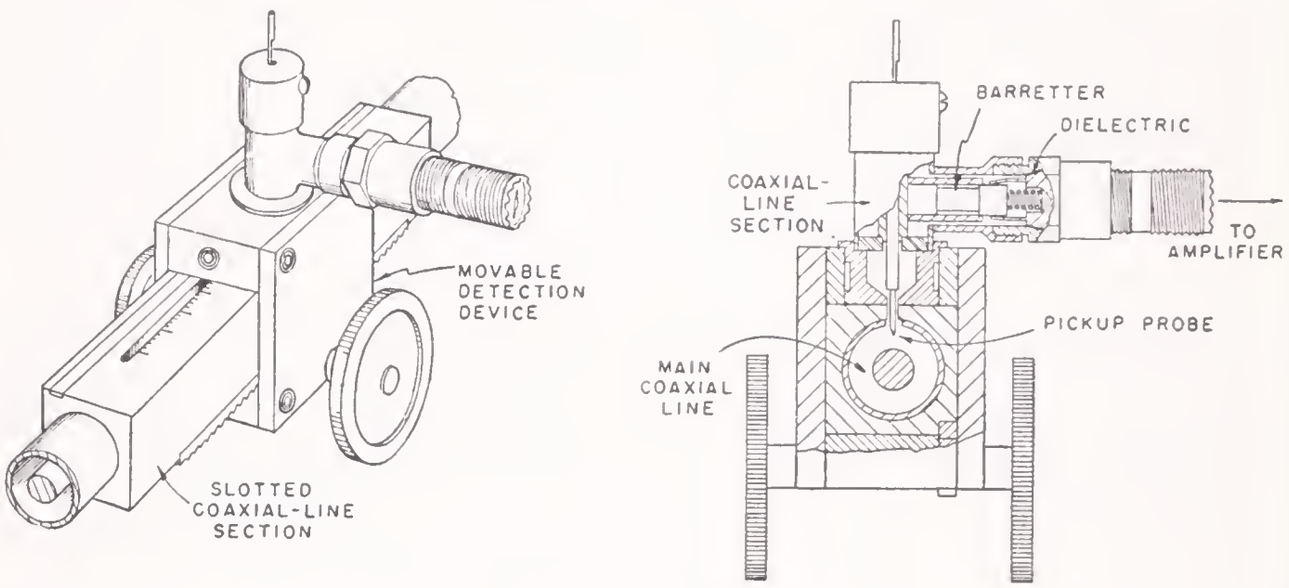


FIG. 60. Slotted-line section for standing-wave measurements.

conductors is radial, the induced probe voltage is proportional to the total voltage between main-line conductors.

The detector element may be a crystal, or it may be a *barretter*, as indicated in Fig. 60*B*. A barretter consists of a short piece of fine platinum wire mounted in a holder which can be inserted in series with the inner conductor of the output-coupling line. The wire is given sufficient resistance to terminate the line as nearly as possible in a matched load. The barretter, which is a form of bolometer, depends for its operation upon the fact that its resistance changes with the amount of power dissipated (and hence the heat generated) in the element. When used as the detector in a standing-wave indicator in radar transmission lines, the barretter is supplied with a small quiescent direct current. The thermal time constant is made small enough so that, when the pulsed r-f power picked up by the coupling probe is dissipated in the barretter, the quiescent direct current is modulated at the pulse repetition frequency. The modulated signal may then be applied to an a-c amplifier, and the ampli-

fied signal used to provide an indication of the main-line voltage. A convenient type of indicator consists of a rectifier in series with a microammeter. The rectified current is then a measure of the main-line voltage.

A machine similar to that shown in Fig. 60 may be used for measuring the standing-wave ratio in rectangular waveguides propagating dominant waves. The slot is cut longitudinally at the center of a broad wall of the guide. Loss of power because of radiation through the slot is thus minimized, since current flow at the slot location is wholly in the longitudinal direction.

Wavemeters. The standing-wave machine of Fig. 60 can also be used to determine the wavelength of r-f signals being propagated in the r-f trans-

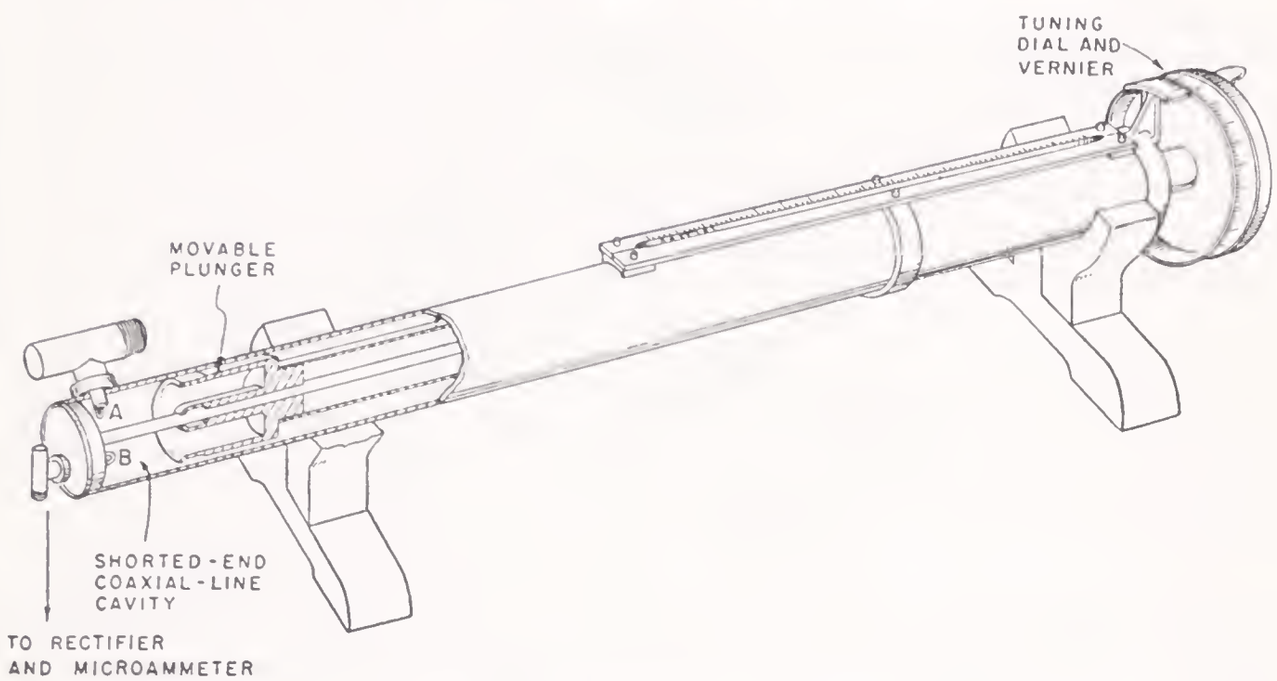


FIG. 61. Coaxial-line wavemeter.

mission system. The wavelength is twice the distance between adjacent minimum-voltage (or maximum-voltage) points. This method of measuring wavelength is not applicable, of course, when the standing-wave ratio is unity and becomes increasingly inaccurate as the value of unity is approached because of the difficulty of locating minimum-voltage points.

A more accurate microwave wavemeter employing a shorted-end coaxial-line cavity is illustrated in Fig. 61. The r-f power is coupled into the cavity by coupling loop *A*, and the indicating device, consisting of a crystal rectifier and microammeter, is supplied with a small amount of power by coupling loop *B*. One end wall of the cavity can be moved by means of the tuning dial located at the end of the instrument.

In order to determine the wavelength of the r-f input signal, the movable plunger is placed at points which yield maximum output-meter readings. These points correspond to multiple half-wave resonances of

the coaxial-line section (Art. 15, Chap. IX). The corresponding plunger positions may be determined accurately from readings of a scale atop the instrument and the vernier scale on the tuning dial. The distance between successive plunger positions yielding maximum output readings is equal to a half wavelength of the input signal. Because the Q of the wavemeter is high, the resonances are well defined and an accurate indication of wavelength is possible.

Echo Box. A piece of equipment which is useful for a variety of test purposes is the echo box, consisting of a cylindrical resonant cavity excited so that resonance is in one of the low-attenuation ($TE_{0,1,n}$) modes (Art. 5, Chap. IX). One of the cavity end walls is built as a movable plunger so that the cavity may be tuned over a range of frequencies.

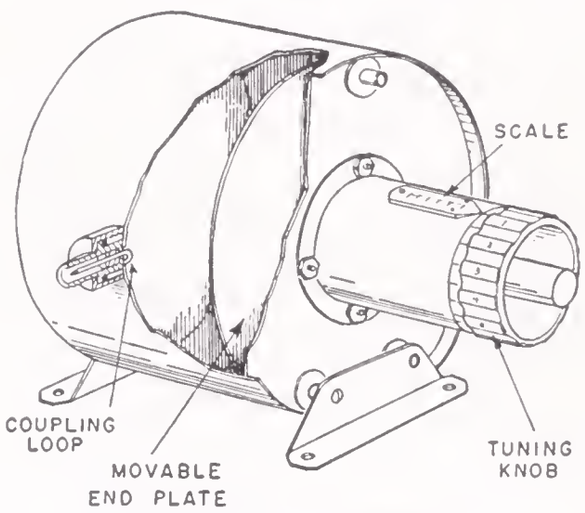


FIG. 62. A 10-cm echo box.

An example of an echo box intended for 10-cm operation is given in Fig. 62. The loop at the cavity periphery serves to couple r-f signals into the cavity, and the knob at the end of the instrument enables the cavity to be tuned to the frequency of the input signal. An indication of resonance is given by a microammeter which responds to the rectified current through a crystal detector attached to an output-coupling loop located on the opposite side of the instrument from the input loop. The frequency of resonance is obtained from a calibration chart which gives the resonant frequency of the cavity in terms of the reading of a dial attached to the tuning knob.

When used to measure the operating frequency of a radar system, the echo box may be coupled to the r-f system through a directional coupler, or a pickup antenna, placed in the vicinity of the radar set, may be used to intercept a small amount of r-f power being radiated. The r-f signals at the pickup antenna are then conducted to the echo box over a short section of coaxial line, and the echo box is tuned to resonance at the frequency of the input signal. Resonance is indicated by a maximum reading of the output microammeter.

A typical 10-cm echo box has a bandwidth of the order of 100 kcps, corresponding to a Q of 30,000, whereas the Fourier components of r-f pulses of radar systems usually encompass a band of the order of several megacycles per second on either side of the carrier frequency. Hence if the echo box is tuned to various frequencies within the spectrum of the r-f pulse, the output-meter reading is a measure of the r-f power contained within a 100-kcps band of frequencies centered at the frequency to which the echo box is tuned. A plot of the echo-box output reading as a function of frequency thus indicates the manner in which the r-f output power of the radar transmitter is distributed in the frequency spectrum of the transmitter pulse. Examples of good and poor spectra of magnetron transmitters and their effects upon the performance of the radar system are given in Art. 11, Chap. XI.

When used to measure the spectrum of an r-f pulse, the echo box has the advantages of simplicity and portability. Its disadvantage is that only one point of the power-spectrum curve can be obtained at a time; the over-all curve must be constructed by plotting echo-box meter readings at a number of different frequencies. (A spectrum analyzer that displays the entire power-spectrum curve on a cathode-ray-tube screen is described in a later paragraph.)

An echo box may also be used to evaluate the over-all performance of a radar system. Over-all performance may be defined as the ratio of P_t, the transmitter power output within the receiver bandwidth as measured at the antenna, to $P_{r\min}$, the minimum echo-signal power required to produce an indication on the cathode-ray tube. That is,

$$\text{Over-all performance} = \frac{P_t}{P_{r\min}} = \frac{\text{power output in the receiver bandwidth}}{\text{minimum detectable receiver power}} \tag{6}$$

The over-all performance may be measured by connecting the echo box to a directional coupler attached to the r-f transmission system. During a transmitter pulse, energy is supplied to the echo box by way of the directional coupler. Oscillations build up in the echo box, and the stored energy increases during the interval T, which represents the pulse duration. During the off time of the transmitter, the echo box returns its stored energy to the directional coupler, which now directs the returned power to the radar receiver. The returned power thus appears on the cathode-ray-tube screen as an echo signal. The energy stored in the echo box at the end of the transmitted pulse decreases gradually, so that a short time after the transmitted pulse the returned power becomes so small that the receiver can no longer detect it. The time between the

termination of the transmitted pulse and the instant at which the returned power falls below the minimum required to produce an indication is called the *ringing time*. The ringing time is a measure of the power output of the transmitter, the efficiency of the r-f system, and the sensitivity of the receiver, since the maximum energy stored in the echo box depends on the power output of the transmitter while the instant at which the receiver will no longer detect the returned signal depends on the losses in the r-f system and the receiver sensitivity. The indication which the echo box gives on the various types of presentations is shown in Fig. 63. The range corresponding to ringing time is marked r_t.

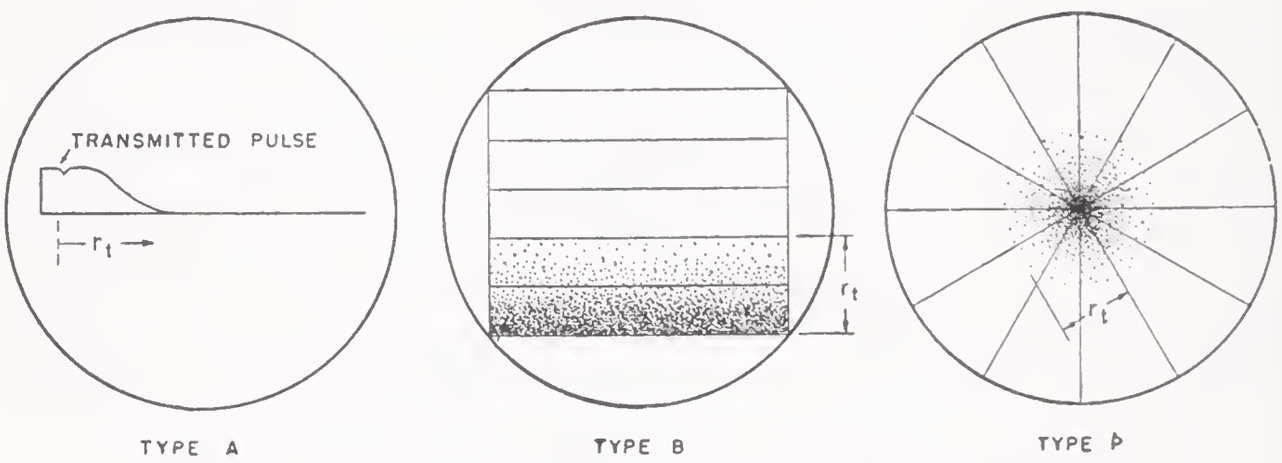


FIG. 63. Echo-box indications on various types of presentations.

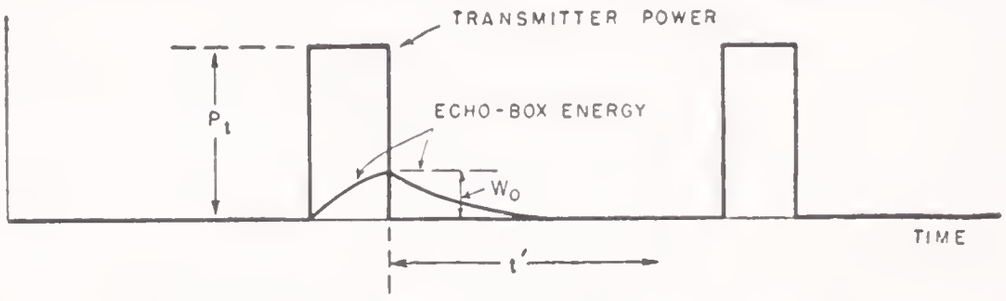


FIG. 64. Power received and returned by echo box.

During a transmitter pulse, the energy stored in the echo box increases exponentially (see Fig. 64). For a given size echo box the stored energy at the end of a transmitter pulse W_0 is proportional to the transmitter power P_t, the pulse duration, the directional-coupler attenuation, and the amount of coupling provided by the echo-box coupling loop. If all factors except P_t are constant, $W_0 = kP_t$. At the end of a transmitter pulse, the echo-box energy is returned to the receiver as a decaying oscillation, the initial rate of energy decay, P_i, being proportional to W_0. Thus the power at the receiver P_r at any instant t' after the transmitter pulse is

$$P_r = k_1 P_i \epsilon^{-t'/T} = k_1 k' W_0 \epsilon^{-t'/T} = k_1 k' k P_t \epsilon^{-t'/T} = k_2 P_t \epsilon^{-t'/T} \qquad (7)$$

where k_1 accounts for the attenuation between echo box and receiver and T is the time constant of the exponential decay of oscillations in the echo box.

From Eq. 6 the relation among the ringing time, the transmitted power, and the minimum detectable power may be determined. If $P_{r_{min}}$ is the minimum power which will produce an indication on the cathode-ray-tube screen and $t_{r'}$ is the corresponding ringing time,

$$P_{r_{min}} = k_2 P_t \epsilon^{-t_{r'}/T} \tag{8}$$

and the ringing time is

$$t_{r'} = T \ln k_2 \frac{P_t}{P_{r_{min}}} \tag{9}$$

Thus the ringing time is proportional to the logarithm of the over-all performance of a radar system. If the over-all performance is expressed in

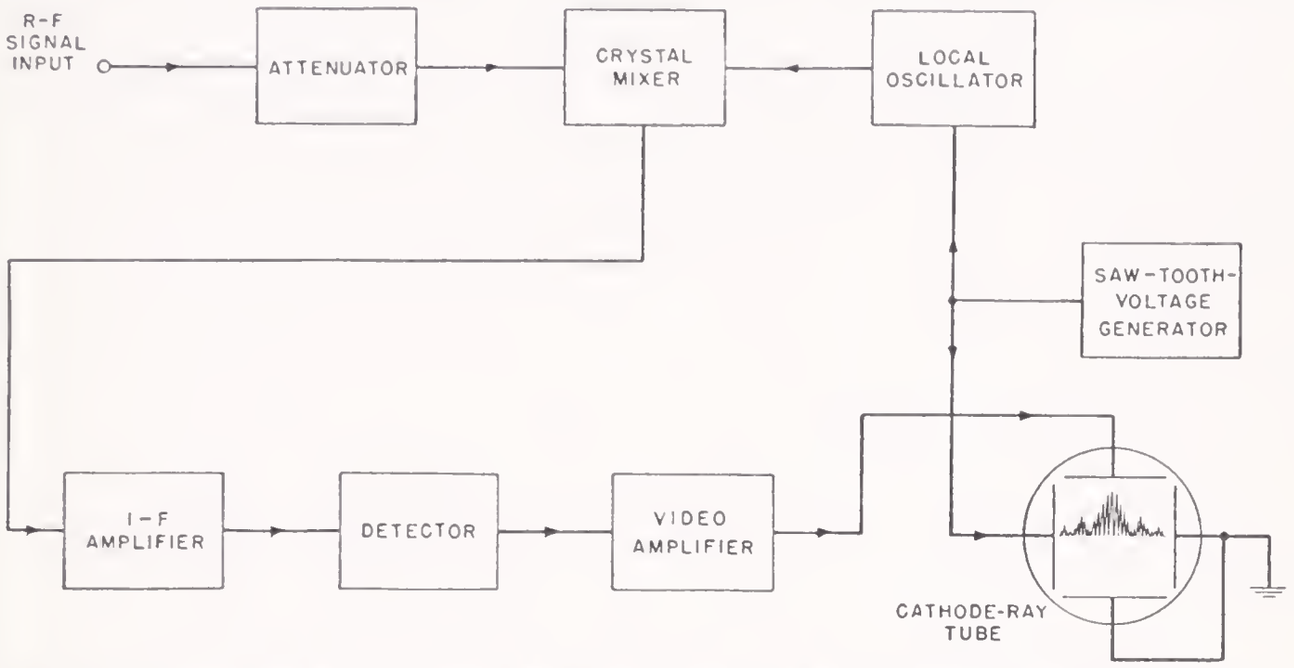


FIG. 65. Block diagram showing basic components of a microwave electronic spectrum analyzer.

decibels, it is directly proportional to the ringing time. An increase in the ringing time due to either an increase in receiver sensitivity or an increase in the power output of the transmitter corresponds to an increase in over-all system performance, which in turn means an improvement in the theoretical maximum range of the system.

Spectrum Analyzer. Although echo boxes may be used to analyze the spectra of r-f pulses, the term spectrum analyzer usually refers to an instrument which displays the entire spectrum on the screen of a cathode-ray tube. A spectrum analyzer is basically a narrow-band superheterodyne receiver with provision for varying the local-oscillator frequency linearly with time. The output signals from the receiver are applied to the vertical-deflection plates of a cathode-ray tube, and the sweep voltage that serves to vary the local-oscillator frequency is applied to the horizontal plates. Because the i-f amplifier bandwidth is narrow—about 100 kcps—the receiver output signal for a given value of local-oscillator fre-

quency is a measure of the power in a very small portion of the input-pulse spectrum. By varying the local-oscillator frequency slowly and moving the cathode-ray-tube beam across the screen in correspondence, the distribution of power in the entire spectrum of the input pulse may be obtained.

The basic components of a spectrum analyzer are shown in block-diagram form in Fig. 65. The r-f input pulses are applied, through an attenuator, to a crystal mixer simultaneously with a signal from the local oscillator. The local oscillator is usually a reflex klystron. A saw-tooth voltage applied to the repeller of the klystron serves to vary the local-oscillator frequency approximately linearly with time. The same saw-tooth voltage also serves as the horizontal-deflection voltage of the cathode-ray tube. The mixer output signals are applied to the remaining portion of the receiver, consisting of a narrow-band i-f amplifier, a detector, and video amplifier. The appearance of the input-pulse spectrum is similar to that shown on the face of the cathode-ray-tube screen in Fig. 65.

CHAPTER XIII

ANTENNAS

The characteristics and operation of any radar antenna system affect the performance of the radar equipment in several ways. The two most important ways pertain to the operation of the antenna as a directional radiator or absorber of radio-frequency energy and to the behavior of the antenna as a circuit element which has to be matched to its associated transmission line.

The directional features of radar antennas are usually their most important characteristics. A directive transmitting antenna makes it possible to illuminate only certain selected targets at any one time, and a directive receiving antenna makes it possible for the radar set to look only in selected directions at a given time. Directivity in both antennas greatly improves the over-all directivity of the system. This is accomplished by illuminating only a few targets at any one time and simultaneously looking directly at these, with the result that perceptible echoes are received only from the chosen targets. A highly directive antenna is also desirable because it allows all the radiated energy to be concentrated in one general direction, and this is enormously helpful from the standpoint of the power required to produce a discernible echo signal.

Because the proper operation of a radar system depends in a fundamental way on the antenna, it is desirable to consider a few of the principles by which antenna performance can be explained. No attempt is made in this chapter to obtain results through the use of electromagnetic theory, but rather the intent is to select those results which are most useful in treating antennas and to show how they may be applied. In general, the discussion to follow will be concerned with transmitting antennas, not because they are more important than receiving antennas, but because the characteristics of a given antenna are the same, if properly interpreted, when the antenna is used for either purpose. This equivalence follows from the reciprocity law, which applies in antenna work and which is an extension of the well-known reciprocity theorem used in electric-circuit analysis.

1. The Half-wave Antenna. Because half-wave antennas are the fundamental radiating elements in most radar arrays, they are an appropriate starting point in a discussion of radar antennas. Physically, a half-wave antenna is a conductor of essentially uniform diameter and

approximately a half wavelength long. Electrically, a half-wave antenna may be regarded as a kind of resonant transmission-line system involving oscillating distributions of charge, current, and voltage along the length of the antenna and electric and magnetic fields in the region around the antenna. The electric and magnetic fields in the region close to the antenna have approximately the configurations shown in Fig. 1. The magnetic-flux lines are circles lying in planes perpendicular to the antenna, with centers on the axis of the antenna. The electric-flux lines lie in planes which contain the antenna itself.

The distribution of charge density, current, and voltage along the antenna at several specific instants during the cycle of oscillation is shown

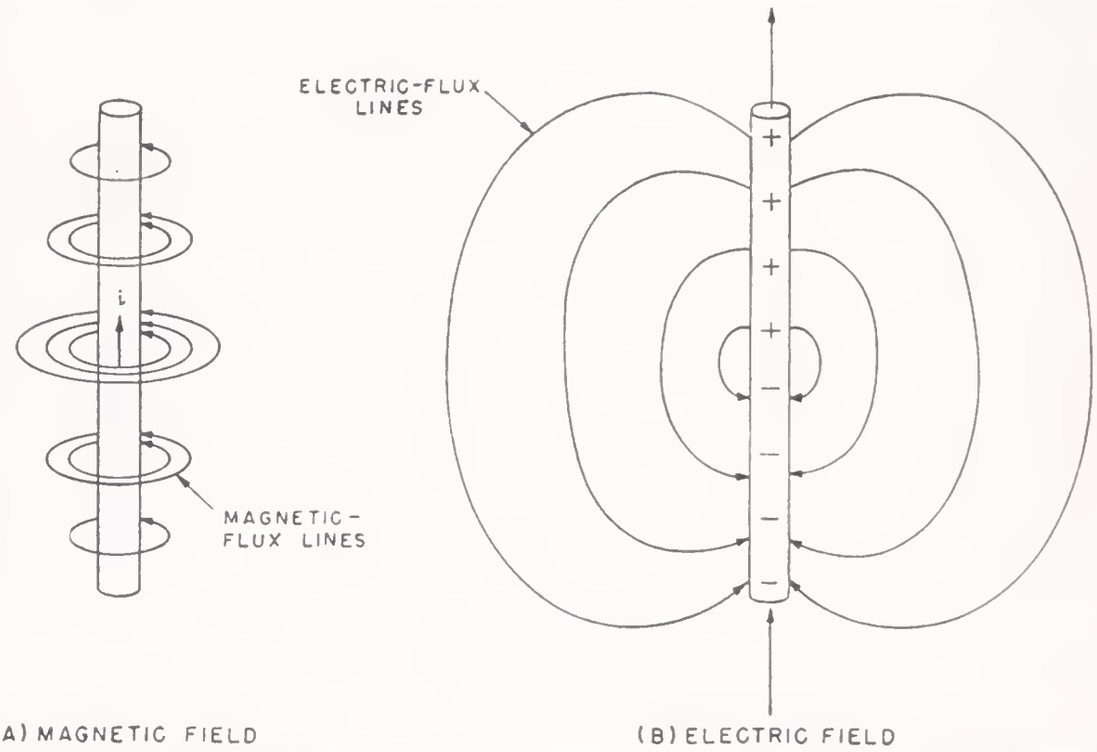


Fig. 1. Approximate configuration of electric and magnetic fields close to the antenna at an instant during the cycle of oscillation when the current flow is upward and the upper part of the antenna is at a positive potential with respect to the lower part.

in Fig. 2. In Fig. 2*A*, the instantaneous distribution of charge density (charge per unit length q/d) is drawn at one-eighth-cycle intervals for a half cycle of the oscillation. Positive values indicate positive charge per unit length of antenna, and negative values indicate negative charge per unit length of antenna. Figures 2*B* and *C* show, respectively, the instantaneous current and voltage distribution at the same instants of the oscillatory cycle. Current flow to the right is plotted positive, and current flow to the left is plotted negative. The voltages of Fig. 2*C* are measured from the center of the antenna; the plots are positive at points which have instantaneous potentials above the potential of the center and negative at points which have instantaneous potentials below the potential at the center. At any point along the antenna the charge, current, and voltage vary sinusoidally with time.

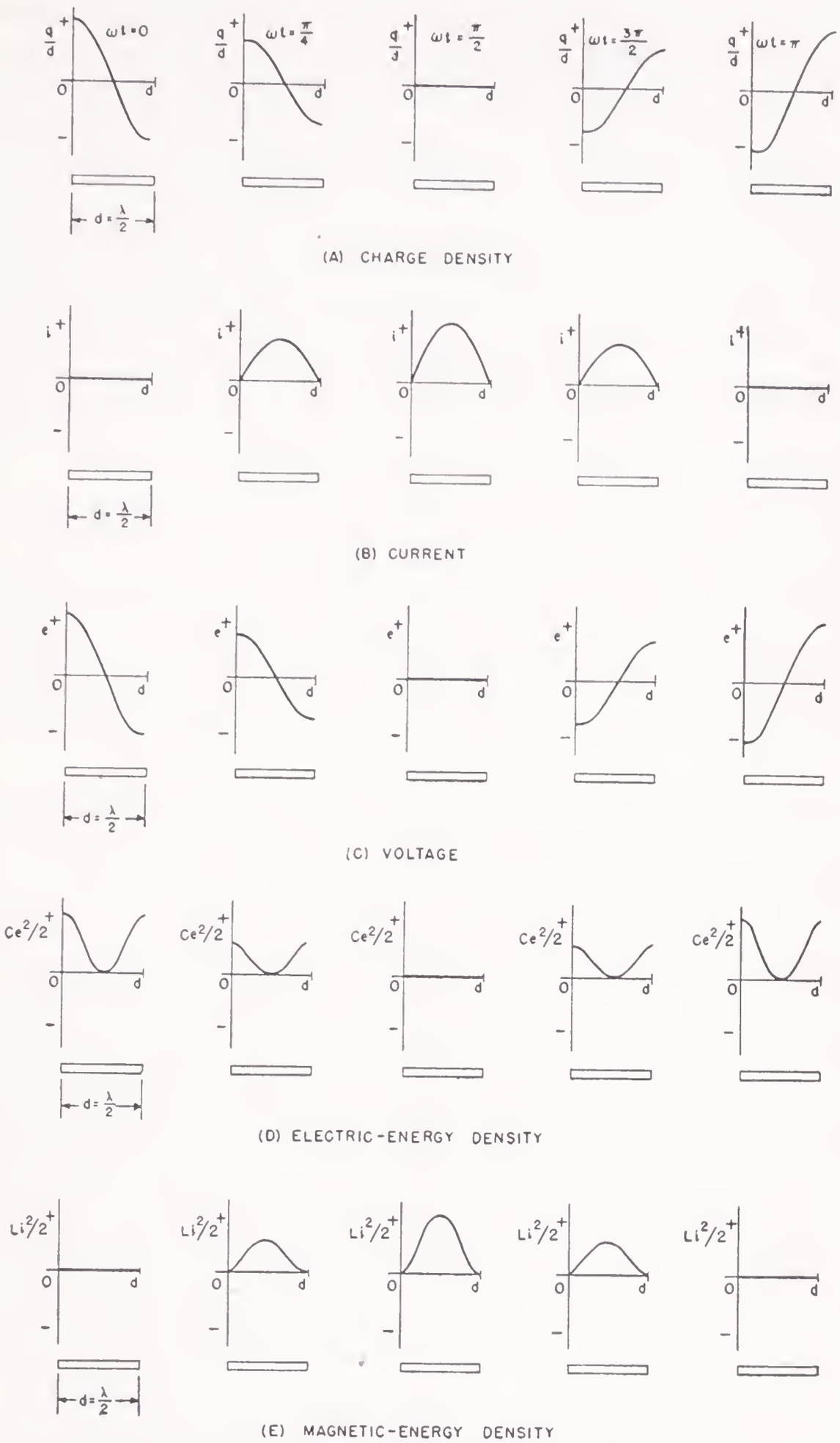


FIG. 2. Distribution of charge per unit length, current, voltage, and energy per unit length along a half-wave antenna at one-eighth-cycle intervals of ωt.

The electric and magnetic fields around the antenna represent stored energy. At any instant the energy stored per unit volume in space is $\frac{1}{2}\varepsilon_0 \mathcal{E}^2$ for the electric field and $\frac{1}{2}\mu_0 \mathcal{H}^2$ for the magnetic field, where ε_0 and μ_0 are, respectively, the permittivity and the permeability of space and $\mathcal{E}$ and $\mathcal{H}$ are the instantaneous values of electric-field intensity and magnetic-field intensity, respectively. The stored energy may also be expressed in terms of the capacitance between corresponding elements at opposite ends of the antenna and the inductance associated with each element of the antenna. The stored energy per unit length of antenna is $\frac{1}{2}Ce^2$ for the electric field, and $\frac{1}{2}Li^2$ for the magnetic field, where C and L are capacitance and inductance per unit length. Although C and L are not constant along the length of the antenna, the energy density (energy per unit length) distributions are approximately the square of the voltage and current distributions, as shown in Figs. 2*D* and *E*.

The areas under the curves of energy per unit length represent the energy stored for the complete length of antenna. It is important to note that, for a certain length of antenna—called a resonant length—which is slightly less than $\lambda/2$, the total stored energy is constant through the cycle. At the instants of maximum current the energy is all stored in magnetic form, and at the instants of maximum voltage it is all stored in electric form, but at all other instants the energy is stored partly in each form.

In addition to the stored energy there is dissipated energy, some being lost as heat in the antenna conductor and some being radiated. In practical antennas the amount of energy lost as heat is negligible compared with that radiated, and the amount radiated per cycle is smaller than the total amount stored.

From an energy viewpoint, a half-wave antenna is very similar to a resonant circuit of moderate Q. In both resonant systems the stored energy exceeds that lost per cycle, and in both the stored energy surges back and forth from electric to magnetic form, the total amount stored remaining constant during the cycle. In both the antenna and the L-C circuit the currents and voltages are predominantly determined by the surging energy and are only slightly affected by the energy furnished by the source of excitation. But the input impedance in both—the impedance presented to the source of excitation—depends upon the power supplied by the source and not upon the surging power. Since the source supplies only the dissipated power, the input impedance is a resistance of magnitude E^2/P, where P is the power supplied and E is the voltage at which it is supplied. When E is large (as in parallel resonance), the resistance is high, and when E is small (as in series resonance), the resistance is low.

A further comparison of half-wave antennas with resonant circuits is

made in Fig. 3, where three methods of exciting or feeding antennas are shown together with the corresponding lumped-parameter analogue. The method shown at the left in Fig. 3 gives an input impedance about equal to the characteristic impedance of a coaxial line and is therefore suitable when a single antenna is to be excited. The method shown at the right in Fig. 3 has an input impedance of several thousands of ohms and is suitable for exciting many half-wave antennas which are effectively in parallel. The method shown in the center diagram gives input impedances intermediate between the other two and is suitable for exciting a few antennas which are effectively in parallel.

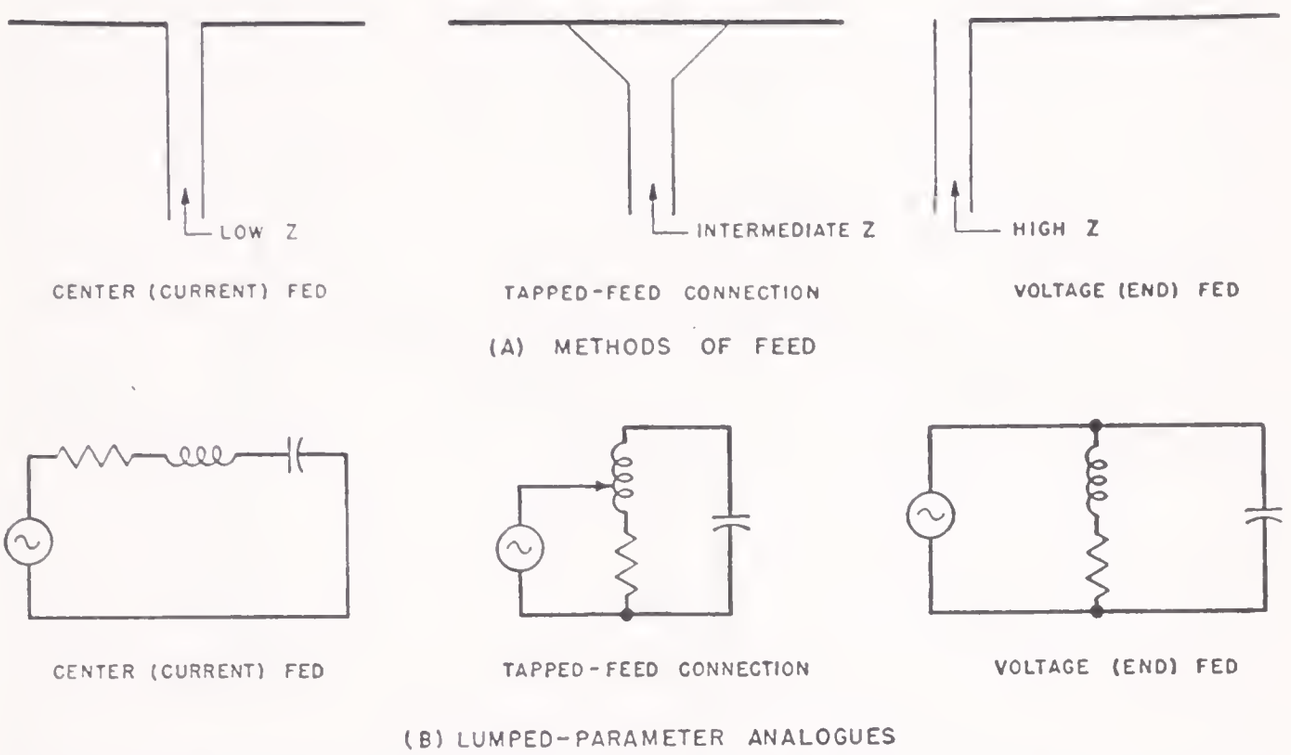


FIG. 3. Three methods of exciting half-wave antennas, and their analogues in lumped parameter resonant circuits.

2. Input Impedances of Center-fed Antennas of Arbitrary Length. Figure 4 shows the input reactance and Fig. 5 the input resistance of a center-fed antenna of arbitrary length. The input impedance is a series combination of the two components. The important regions of the curves correspond to antenna half lengths near $\lambda/4$ and near $\lambda/2$. The former represents a center-fed half-wave antenna, whereas the latter represents a pair of end-fed half-wave antennas excited in phase. The half length of the antenna is used in plotting in order to show the resemblances between the antenna-reactance curves and the input-reactance curves for an open-ended transmission line.

In the regions of principal interest the reactance curves of Fig. 4 are nearly straight lines, the slopes of which depend on the diameter of the antennas expressed in wavelengths. The slopes of the reactance curves decrease as the antenna diameter increases. This characteristic is impor-

tant in radar antennas which need to be insensitive to small changes in frequency. The curves show that antennas of large diameter present less than a specified amount of reactance, say 1 ohm, over a greater range of antenna length than slender antennas do. Thus a given length of antenna has less than 1 ohm of reactance over a wider range of frequency when the antenna has a large diameter than when it has a small diameter. Radar antennas are commonly made of tubing and frequently have diameters in excess of $\lambda/20$.

Figure 5 shows that the input resistance also depends on antenna diameter. This dependence is more pronounced when the half length is about

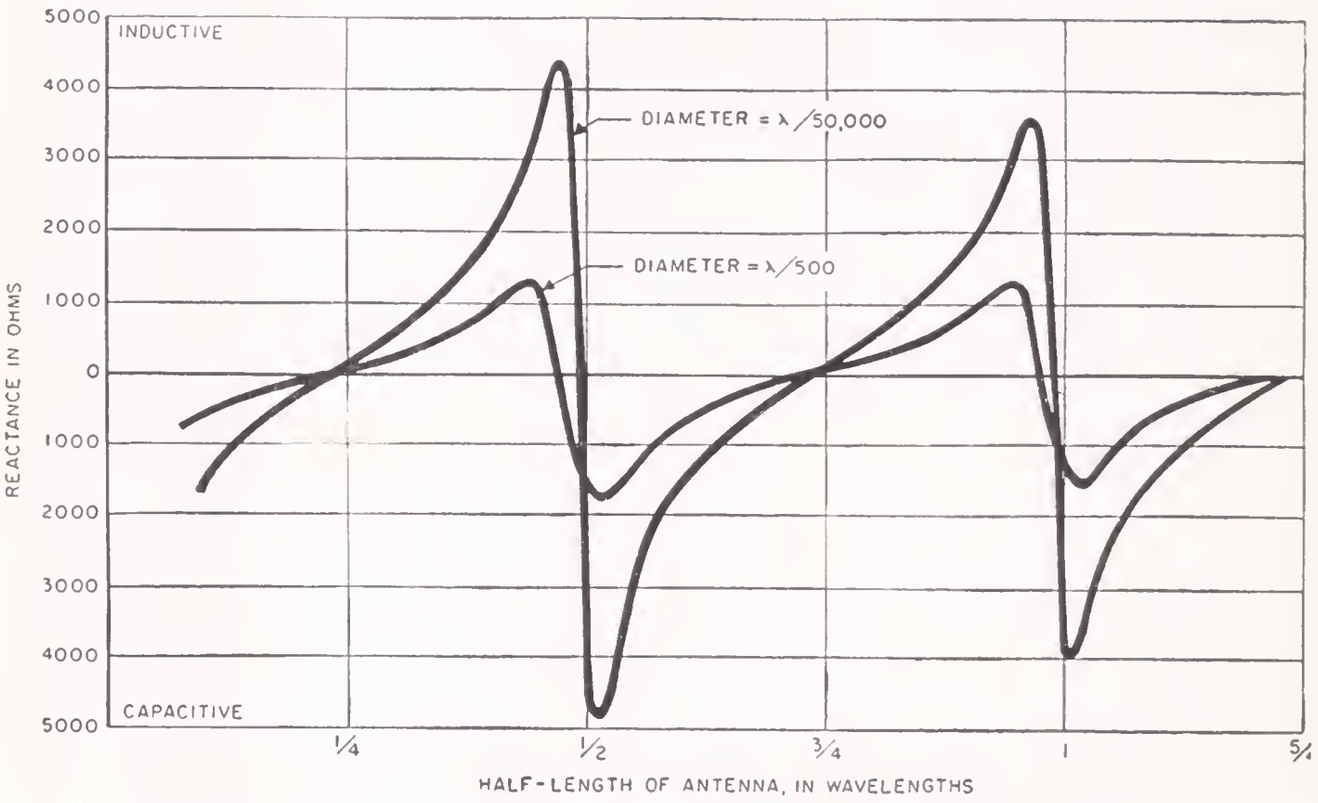


FIG. 4. Reactance at input terminals of a center-fed antenna of arbitrary length.

$\lambda/2$ than when the half length approximates $\lambda/4$, as it does for a single center-fed antenna. The values for the single half-wave antenna are not readable on the curve, but the component representing radiation ranges from 73 ohms for infinitely thin antennas through 64 ohms for a diameter of 0.0001λ, 55 ohms for a diameter of 0.01λ, to less than 50 ohms for certain large-diameter radar antennas.

The input resistance of a center-fed antenna comprises radiation resistance and the effective resistance of the antenna conductors at the operating frequency. Because antennas usually employ short, thick elements, the latter component is negligible. Radiation resistance may be defined as the ratio of the power radiated from the antenna to the square of the current at the input terminals. Sometimes radiation resistance is defined in terms of the maximum current of a half-wave antenna rather than in terms of the current at the input terminals.

The input impedance of a center-fed antenna is the complex number that results from adding the resistance and reactance obtained from the curves of Figs. 4 and 5. It should be borne in mind that these curves are valid only when the antenna is isolated from other current-carrying elements. The presence of additional elements, such as parasitic ele-

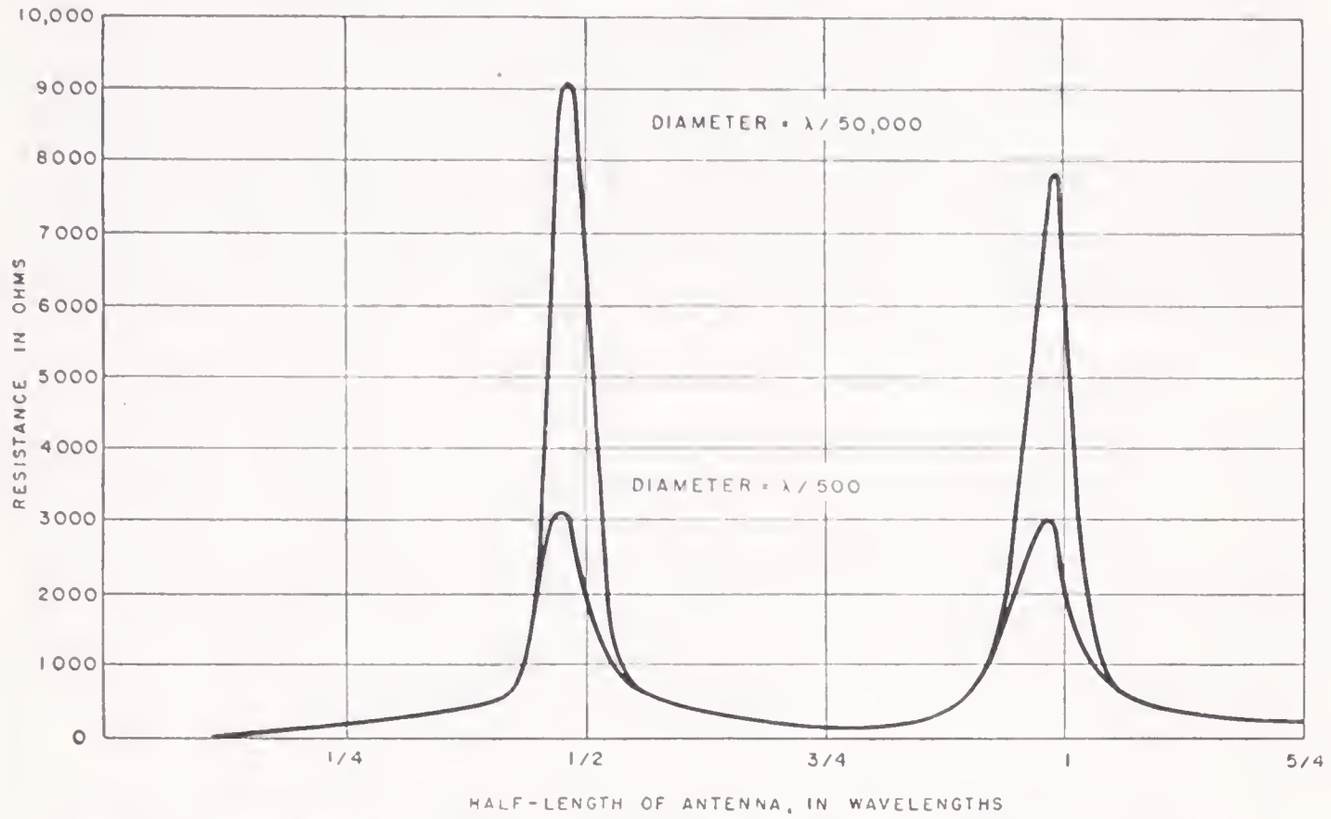


FIG. 5. Resistance at input terminals of a center-fed antenna of arbitrary length.

ments, reflecting screens, or adjacent antennas in an array, modifies the input impedance of the antenna because of mutual interaction of the elements. Since the flow of current in any one element induces a voltage in all the others, each radiator is coupled to each of the other elements.

A feature of Fig. 4 which is not easily discernible is that the over-all lengths at which the reactance is zero are less than $\lambda/2$ and λ. The amount by which an antenna with zero reactance is shorter than these lengths depends on the antenna diameter. For very slender antennas the shortening is slight, but for large-diameter antennas or for special shapes, such as the noncylindrical antenna of Fig. 6, the resonant length may be as much as 20 per cent shorter than $\lambda/2$. Special shapes, such as the one shown in Fig. 6, have the advantage of being insensitive to small changes in frequency and at the same time are not so subject to corona (breakdown of the air because of large potential gradients) as cylindrical antennas are.

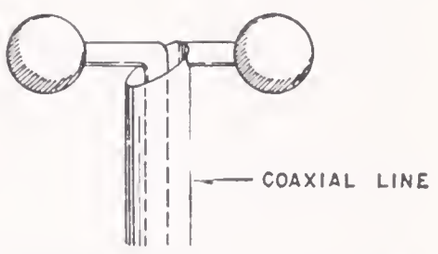


FIG. 6. Noncylindrical half-wave antenna.

It is of interest to consider the reactance curves of Fig. 4 in terms of

stored energy. When an antenna is of resonant length, the stored energy is constant during the cycle and the source supplies only the dissipated energy. To the source the antenna appears, therefore, as a purely resistive load which requires real power but no reactive power. If the antenna is center fed and slightly less than the resonant length, the stored energy is not constant during the cycle. This can be seen by taking a strip from the center of the energy-density curves of Fig. 2, the width of the strip being equal to the reduction in antenna length. The shortened antenna stores more energy in electrical form at the instant of maximum voltage than it stores in magnetic form at the instant of maximum current. Consequently the source must supply capacitive reactive power in addition to real power, and it therefore sees the antenna

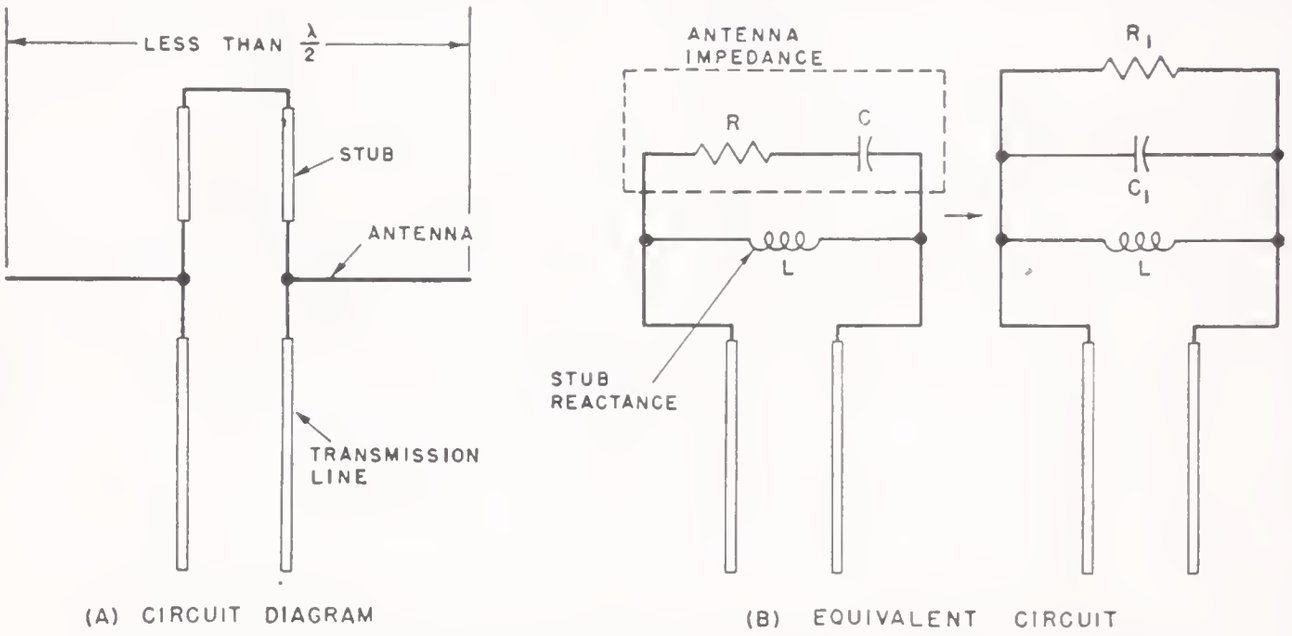


FIG. 7. A method of adjusting the input impedance of a center-fed antenna.

as a capacitive impedance. It may be shown in a similar manner that the reactance of a center-fed antenna slightly longer than $\lambda/2$ is inductive. It can be seen also that the reactance of an antenna of length slightly greater than λ is capacitive and that the reactance of an antenna of length slightly less than λ is inductive, as indicated in Fig. 4.

In radar it is frequently necessary to adjust the input impedance of antennas. A common method of impedance changing consists in shortening a center-fed, half-wave antenna so that it presents a capacitive impedance at its input terminals and then adding a shorted-end transmission-line stub less than $\lambda/4$ long in parallel with the input to the antenna. Such an arrangement, with its equivalent circuit, is shown in Fig. 7. At a single frequency the antenna impedance may be represented by a resistance R and capacitance C in series, and the stub by an inductance L in parallel with the antenna impedance. If the series combination of R and C is replaced by equivalent shunt elements R_1 and C_1, as indicated in Fig. 7*B*, and the reactive component canceled by means of

the stub, the input impedance becomes the resistance R_1. A wide variety of values is possible for R_1, and these values depend upon the length to which the antenna is trimmed. In addition to serving as an impedance-changing device, the stub may also be used as a mechanical support for the antenna by attaching the shorting bar rigidly to a supporting frame.

If the antenna conductors are made of hollow tubing, coaxial-line stubs may be inserted in the tubing, as shown in Fig. 8. The inner surfaces of the antenna elements serve as the outer conductors of the stubs. The stubs are in series, and each is made less than one-eighth wave in length. The total reactance of the stubs is therefore inductive and is in parallel with the input impedance of the antenna. The antenna may be supported from ground by connecting a conductor to point A, the electrical center of the antenna.

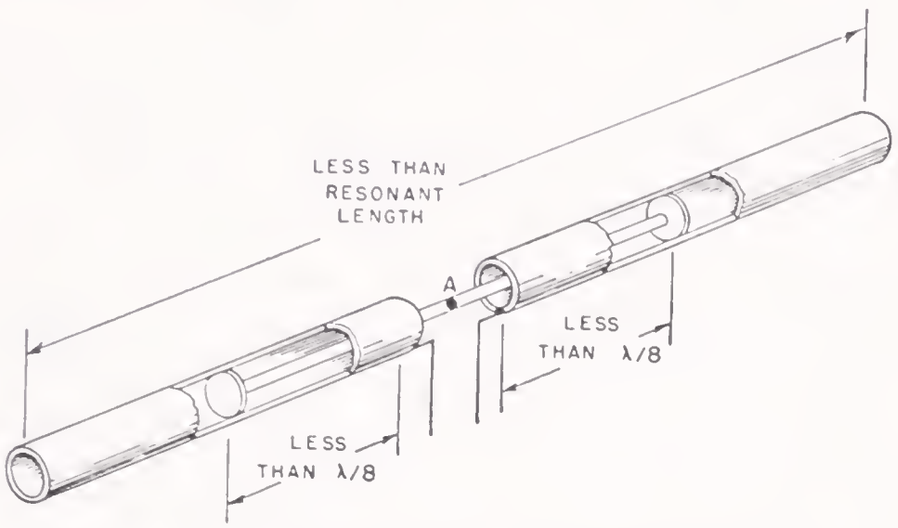


FIG. 8. Center-fed antenna with coaxial stubs.

3. Radiation Fields of a Half-wave Antenna. Thus far only the fields near the antenna have been considered. These fields are associated with the stored energy. Electric and magnetic fields are also associated with the radiated energy, and these are called radiation fields. The radiation fields, which persist to great distances, determine the directional characteristics and input resistance of the antenna, whereas the fields near the antenna are principally responsible for the behavior of the antenna as a resonant-circuit element.

Figure 9 shows the manner in which the radiation fields are propagated away from the antenna. The electric and magnetic fields are represented by separate sets of flux lines which are at right angles to each other and to the radial direction of propagation. The picture is representative of any plane containing the antenna and applies for a single instant of time.

The magnetic-flux lines are circular lines having the axis of the antenna as their axis, so that they appear in Fig. 9 as dots and crosses. At intervals of one-half wavelength along a radial direction, these flux lines reverse direction. This reversal is appropriate because the mag-

netic-flux density varies along the radial direction in approximately a sinusoidal manner, as indicated in the figure.

The electric-flux lines are also closed and consist essentially of arcs of circles lying in planes containing the antenna and joined in the manner shown. They, too, reverse direction at precisely the places where the magnetic-flux lines reverse direction, and their density varies along the

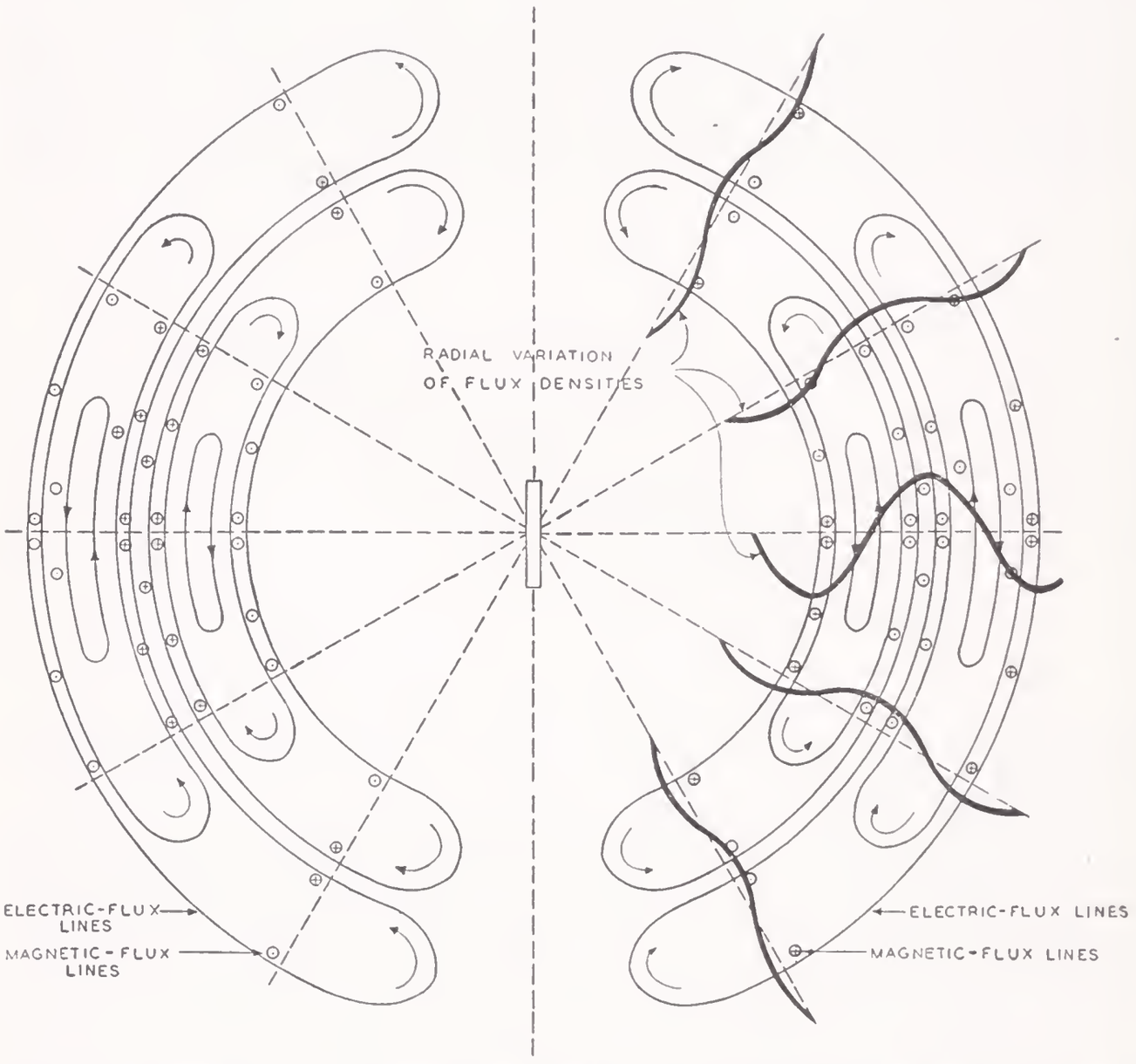


FIG. 9. Configuration of the electric and magnetic fields as they travel radially from a half-wave antenna.

radial direction in the way that the magnetic-flux density varies. In fact, the electric-flux density is everywhere proportional to the magnetic-flux density.

An important feature of Fig. 9 is the way in which the peak flux densities vary with the radial distance (measured from the center of the antenna) and with the angular orientation of the radial direction with respect to the antenna axis. If the radial distance is fixed, the peak electric-flux density is clearly small at the antenna axis, where only a few of the flux lines appear, and rises to a maximum value at an angle of

90 deg from the axis, where all the flux lines appear. If the angular orientation with respect to the axis is fixed and the radial distance is increased, the peak electric-flux density must decrease inversely with the radial distance. This is because the number of flux lines between any adjacent pair of flux zeros is the same, and the area of the circular band of width $\lambda/2$ through which these lines pass increases in proportion to the radial distance.

As stated above, the configuration of flux lines shown in Fig. 9 applies for a single instant of time. As time goes on, the flux lines travel radially at the velocity of light to form a wave motion, and new flux lines are created at the antenna to replace those which journey outward. At any point in space, the electric and magnetic fields oscillate, the oscillation frequency being the same as the frequency of the antenna current. The variations in the magnitudes of electric-field intensity $\mathcal{E}$ and magnetic-field intensity $\mathfrak{IC}$ are in phase. The directions of $\mathcal{E}$ and $\mathfrak{IC}$ in space, as given by the corresponding fluxes in Fig. 9, are at right angles to each other and to the radial direction of propagation.

The foregoing discussion is a description in terms of geometry and flux lines of what takes place as the waves are propagated. In physical terms the mode of propagation is as follows. At every point in space the time-varying magnetic field induces a voltage gradient in space, which is the electric field. This electric field also varies with time, and its variation with time is equivalent to a current, even though it is not associated with a movement of charge. This equivalent current, called the displacement current, establishes a magnetic field in precisely the way that a conduction current does. This magnetic field is the original magnetic field. Thus the varying magnetic field produces a varying electric field, which in turn, through its associated displacement current, sustains the varying magnetic field. Each field supports the other, and neither can be propagated by itself without setting up the other.

4. Analysis of Radiation Fields of a Half-wave Antenna. In the preceding article the fields surrounding a half-wave antenna are described qualitatively. In this article the fields will be described quantitatively at any point in space in terms of both their magnitude and direction. In order to do this, it is first necessary to establish reference directions for the quantities involved.

The first reference direction is that for the antenna current. This, of course, is taken along the antenna in either one of the two possible directions. Antenna current is then specified in the usual way by assigning positive magnitudes to the current when it flows in the reference direction and negative magnitudes when it flows opposite to the reference direction.

In addition to the current reference at the antenna, which can be arbitrarily selected, three reference directions are required at each point

in space, and these are not arbitrary. How they are related to the current reference is conveniently stated in terms of an imaginary spherical surface which surrounds the antenna and passes through the point in question, as shown in Fig. 10. The center of the spherical surface is coincident with the center of the antenna, and the axis of the sphere is taken to be the axis of the antenna. Poles are defined by the inter-

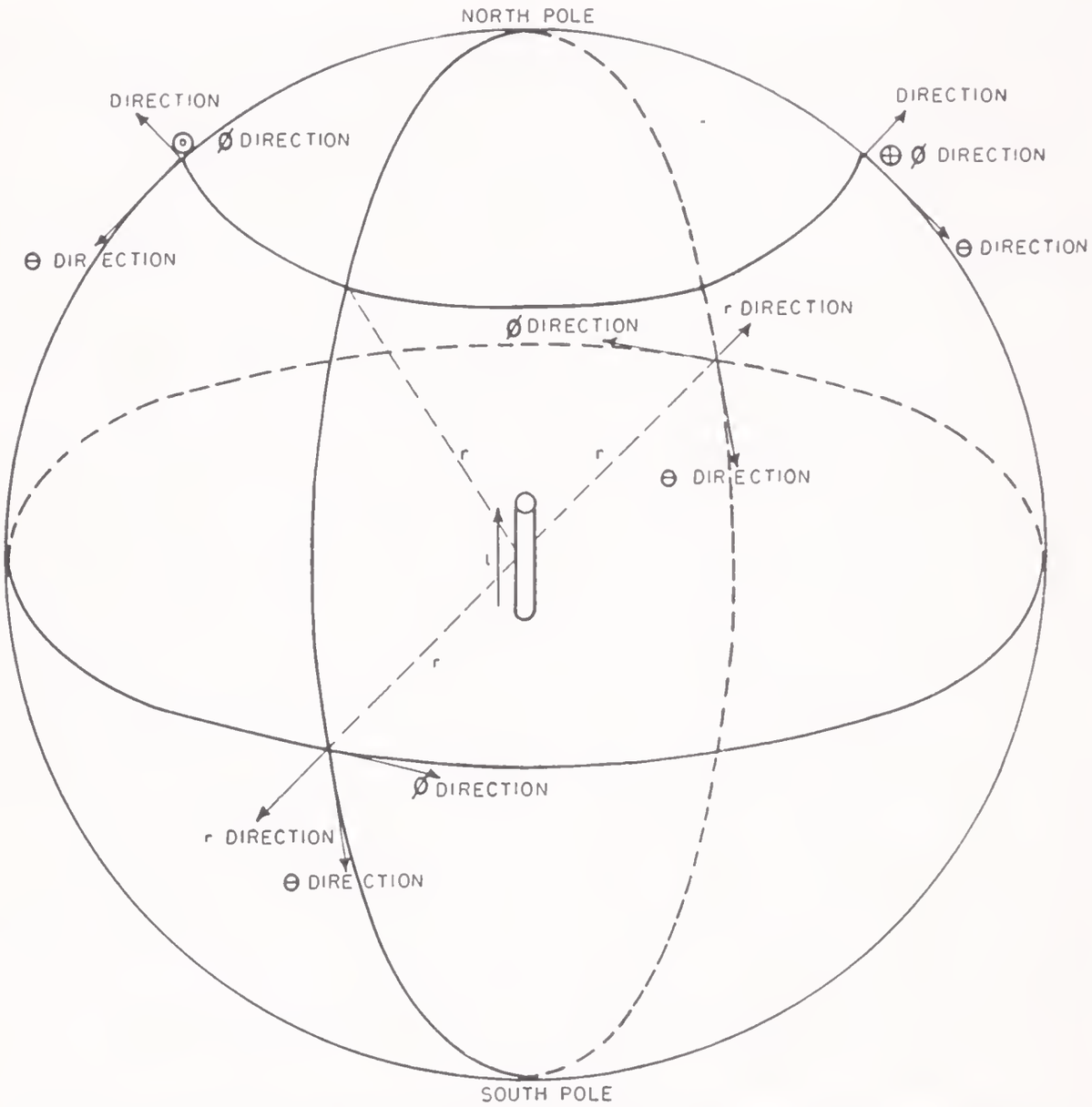


FIG. 10. Coordinates and reference directions used with half-wave antennas.

sections of the axis with the spherical surface, and the one toward which the current reference points is chosen as north. Since the current reference direction in Fig. 10 is upward in the antenna, the north pole appears at the top of the diagram.

At any point on the surface of this sphere three directions are defined as follows: The θ direction is tangent to the surface of the sphere and directed southward on a meridian line, the ϕ direction is tangent to the surface of the sphere and directed eastward on a line of constant latitude, and the r direction is perpendicular to the surface of the sphere and

directed outward. These three directions are mutually perpendicular in space and will serve as the reference directions for describing the fields.

The electric field laid down by a half-wave antenna has a principal component along the θ direction and a small component along the r direction, the latter being necessitated in accordance with Fig. 9 by the joining of the circular arcs. In the discussions which follow, the small radial component will not be considered. To this approximation, the electric field lies wholly in the θ direction (or opposite to it), and the complex number representing the field in this direction will be designated as $\bar{\mathcal{E}}_\theta$. Similarly, the magnetic field lies wholly in the ϕ direction (or

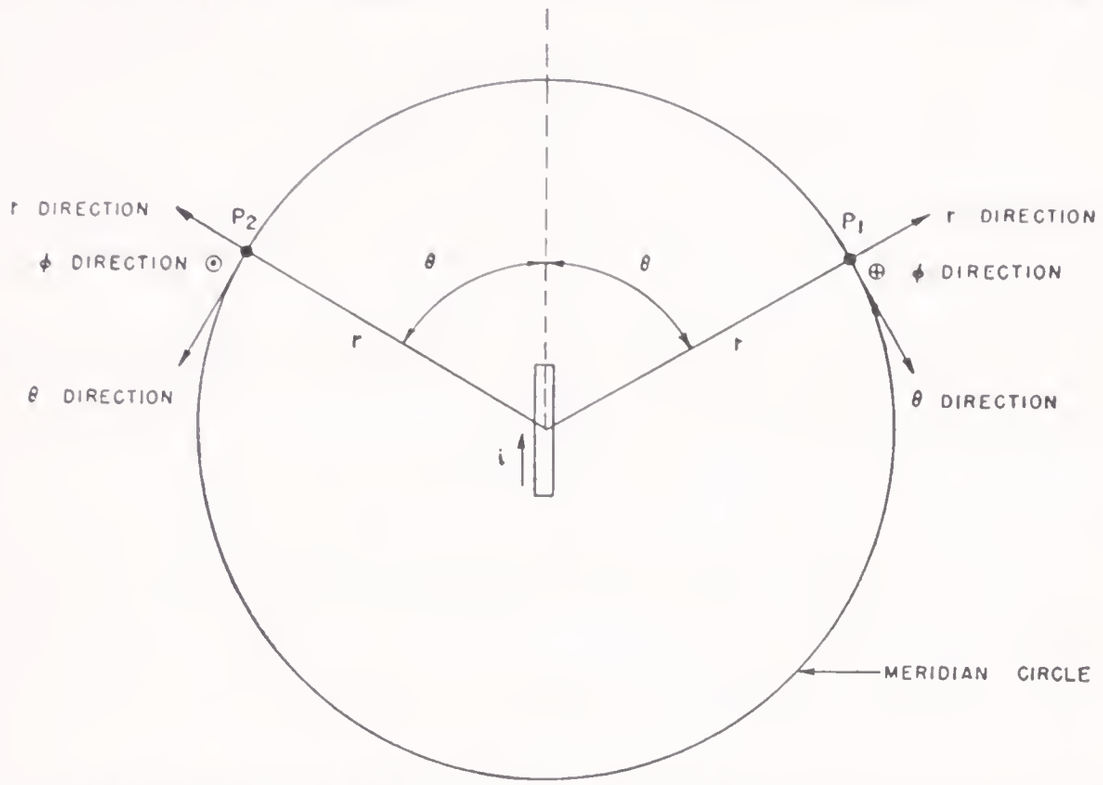


FIG. 11. Plane diagram showing coordinates r and θ and the reference directions at two typical points, P_1 and P_2.

opposite to it), and the complex number representing the magnetic field in this direction will be designated as $\overline{\mathcal{H}}_\phi$. To the approximation that the electric field can be considered to be wholly in the θ direction, the Poynting vector can be considered to be in the r direction. The instantaneous power density in the r direction is designated as $\mathcal{S}_r$.

In order to evaluate $\bar{\mathcal{E}}_\theta$ and $\overline{\mathcal{H}}_\phi$ at any point, it is necessary to introduce the spherical coordinates r and θ. The coordinate r is the radius of the sphere passing through the point in question, that is, the radial distance from the point to the center of the antenna. Angle θ is either the large or the small angle in space between the antenna axis and the radial line drawn from the center of the antenna to the point in question. Because of symmetry it is not necessary in stating magnitudes to distinguish one meridian line from another, so that the spherical angle ϕ is not introduced.

A plane diagram which clearly shows the three directions and the two coordinates so far introduced is shown in Fig. 11. This diagram applies

only at points lying in the plane of the page, but it is representative of any meridian circle. Point P_1 is representative of positions along the right half circle, and P_2 is representative of positions along the left half circle.

Expressions for $\bar{\mathcal{E}}_\theta$ and $\bar{\mathcal{H}}_\phi$ may now be given. Let the current at the center of the antenna be represented by the complex number $I = |I|\epsilon^{j0} = (I_m/\sqrt{2})\epsilon^{j0}$. The complex numbers standing for the electric- and magnetic-field intensities at a point distant r from the antenna are

$$\bar{\mathcal{E}}_\theta = |\bar{\mathcal{E}}_\theta|\epsilon^{-j\left(\frac{360°r}{\lambda} - 90°\right)} = \frac{\mathcal{E}_{\theta m}}{\sqrt{2}}\epsilon^{-j\left(\frac{360°r}{\lambda} - 90°\right)} \tag{1}$$

and

$$\bar{\mathcal{H}}_\phi = |\bar{\mathcal{H}}_\phi|\epsilon^{-j\left(\frac{360°r}{\lambda} - 90°\right)} = \frac{\mathcal{H}_{\phi m}}{\sqrt{2}}\epsilon^{-j\left(\frac{360°r}{\lambda} - 90°\right)} \tag{2}$$

The peak value of the power density $\mathcal{S}_{rm}$ is

$$\mathcal{S}_{rm} = \mathcal{E}_{\theta m}\mathcal{H}_{\phi m} \tag{3}$$

In these expressions $\mathcal{E}_{\theta m}$ is the peak value of the electric field in the θ direction and is related to I_m and r as follows:

$$\mathcal{E}_{\theta m} = \sqrt{2}\,|\bar{\mathcal{E}}_\theta| = \frac{60I_m}{r}F_0(\theta) \tag{4}$$

Also, $\mathcal{H}_{\phi m}$ is the peak value of the magnetic field and is given by the expression

$$\mathcal{H}_{\phi m} = \sqrt{2}\,|\bar{\mathcal{H}}_\theta| = \frac{1}{120\pi}\mathcal{E}_{\theta m} \tag{5}$$

The factor $F_0(\theta)$ has the value

$$F_0(\theta) = \frac{\cos\,(90°\cos\,\theta)}{\sin\,\theta} \tag{6}$$

In the above relations, electric-field intensity is in volts per meter, magnetic-field intensity is in amperes, and power density is in watts per square meter when current is expressed in amperes and r and λ are in meters.

Several important features of Eqs. 1 to 3 may be listed. They are:

1. At a point in space distant r from the antenna, the electric field is delayed in phase $[(360°r/\lambda) - 90°]$ with respect to the current at center of the antenna.

2. At each point of space the electric-field intensity in the θ direction and the magnetic-field intensity in the ϕ direction are in phase with each other.

3. The instantaneous power density $\mathcal{S}_r$ is never negative but pulsates at double frequency between zero and a maximum value of $\mathcal{E}_{\theta m}\mathcal{H}_{\phi m}$ and has an average value of $\mathcal{E}_{\theta m}\mathcal{H}_{\phi m}/2$.

The phase relationships existing among the antenna current and electric- and magnetic-field intensities at a remote point in space may be represented graphically by a complex-number diagram shown in Fig. 12. Antenna current is taken as the reference wave and is the complex number

$$I = (I_m/\sqrt{2})\epsilon^{j0}$$

The complex number $\bar{\mathcal{E}}_\theta$, representing the electric-field intensity in the θ direction at a point distant r from the antenna, has a magnitude $\mathcal{E}_{\theta m}/\sqrt{2}$, where $\mathcal{E}_{\theta m}$ is given by Eq. 4, and is delayed in phase with respect to I by an angle $[(360°r/\lambda) - 90°]$. Thus $\mathcal{E}_\theta$ is retarded with respect to I a whole number of revolutions plus a fraction of a revolution, as indicated in the diagram. The magnetic-field intensity in the ϕ direction at the same point is the complex number

$$\overline{\mathcal{H}}_\phi = \frac{1}{120\pi}\frac{\mathcal{E}_{\theta m}}{\sqrt{2}}\epsilon^{-j\left(360°\frac{r}{\lambda}-90°\right)}$$

and is in phase with $\bar{\mathcal{E}}_\theta$.

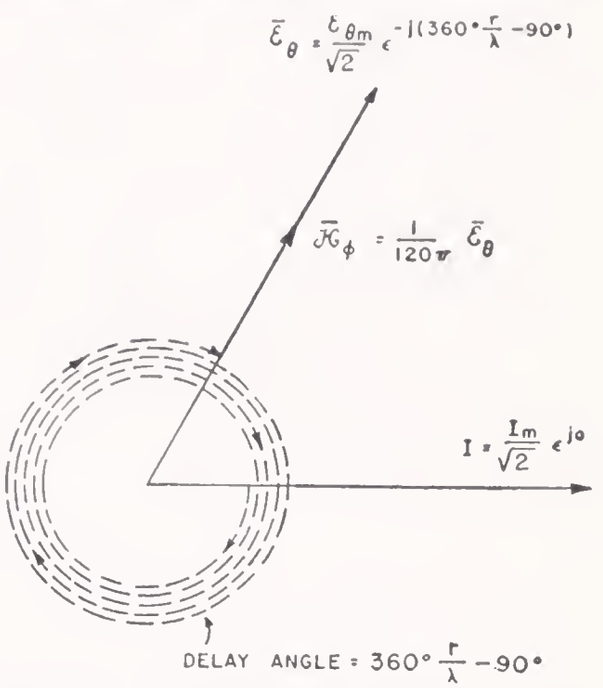

$\bar{\mathcal{E}}_\theta = \frac{\mathcal{E}_{\theta m}}{\sqrt{2}}\epsilon^{-j(360°\frac{r}{\lambda}-90°)}$
$\bar{\mathcal{H}}_\phi = \frac{1}{120\pi}\bar{\mathcal{E}}_\theta$
$I = \frac{I_m}{\sqrt{2}}\epsilon^{j0}$
DELAY ANGLE = $360°\frac{r}{\lambda} - 90°$

FIG. 12. Complex-number diagram showing antenna current and fields at a distance r from the center of a half-wave antenna.

It should be noted that the delay angle of $\bar{\mathcal{E}}_\theta$ is usually enormously large, being the sum of an integer times 360 deg plus an angle less than 360 deg, which is the visible angle on the complex-number diagram. If the location of $\bar{\mathcal{E}}_\theta$ on the diagram is to be known to 1 deg, r must be known to $\lambda/360$ units of r regardless of the size of r. Hence one must exercise extreme care in evaluating r when the phase of $\bar{\mathcal{E}}_\theta$ is being determined. In a calculation of the magnitude of $\bar{\mathcal{E}}_\theta$, however, extremely precise values of r are not necessary, since a fixed percentage error in r causes the same percentage error in the magnitude of $\bar{\mathcal{E}}_\theta$.

Another important feature of the field relations given by Eqs. 1 through 6 is the manner in which $\mathcal{E}_{\theta m}$ depends upon I_m, r, and θ. For a given antenna current and given spherical surface where r has a single value, $\mathcal{E}_{\theta m}$ depends only on θ. The actual relationship of $\mathcal{E}_{\theta m}$ to θ is given by the factor $F_0(\theta)$, defined by Eq. 6. Values of $F_0(\theta)$ are given in Table 1 and are plotted in both rectangular and polar coordinates in Fig. 13.

Because of symmetry about $\theta = 90°$, it is unnecessary to tabulate $F_0(\theta)$ beyond $\theta = 90°$. It is because of this symmetry that the angle θ can be measured from either end of the axis of the antenna. When using

$F_0(\theta)$, it will be convenient to remember that $F_0(\theta)$ is very roughly equal to sin θ where θ does not exceed 180 deg (see Fig. 13*A*).

From Fig. 13 it is evident that as a single meridian line on the spherical surface is traced, the magnitude of the electric field changes from zero at the north pole to a maximum at the equator and back to zero at the south pole. This is true for any meridian line and is consistent with Fig. 9.

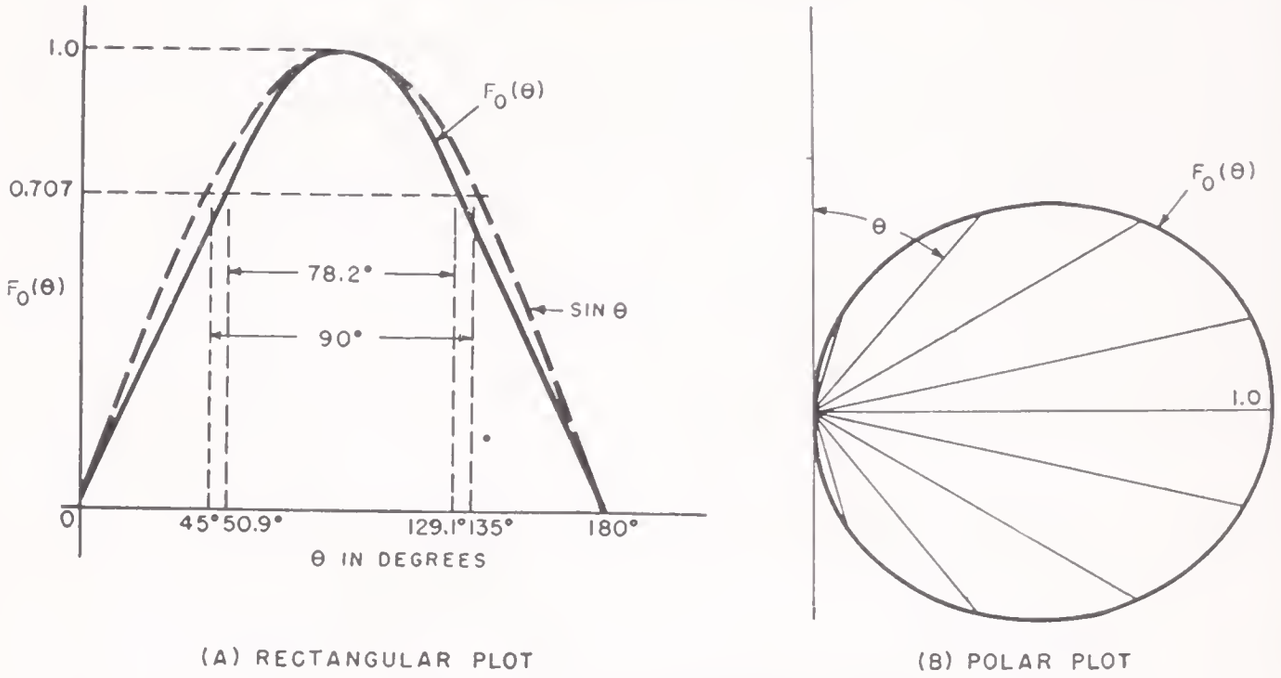


Fig. 13. Plots of the function $F_0(\theta)$.

Table 1. Values of $F_0(\theta)$

θ	$F_0(\theta)$	θ	$F_0(\theta)$	θ	$F_0(\theta)$	θ	$F_0(\theta)$
0°	0	25°	0.346	50°	0.696	75°	0.955
5°	0.080	30°	0.418	55°	0.758	80°	0.978
10°	0.141	35°	0.490	60°	0.817	85°	0.995
15°	0.205	40°	0.557	65°	0.868	90°	1 00
20°	0.275	45°	0.627	70°	0.914	95°	0.995

5. Patterns of a Half-wave Antenna; Pattern Interpretation. Antenna patterns are of two kinds, field-intensity patterns and power-density patterns. The first is a plot of the electric-field intensity at a fixed distance from the antenna as a function of angle measured in the particular plane for which the pattern applies. The second is a plot of the power density at a fixed distance from the antenna against angle in a particular plane. Magnetic-field intensity is not considered because plots of it are the same, except for scale, as those of the electric field. For either kind of pattern the plots can be made for a large number of planes in space, but usually only three principal mutually perpendicular planes are considered. The plots are normally made in polar coordinates; that is, radial lines are drawn from a point (the pole), and the length of each

line is made proportional to the magnitude of the electric-field intensity $\mathcal{E}_{\theta m}$ or power density $\mathcal{S}_{rm}$ at a given distance r in the direction of plotting. The pattern is a curve joining the ends of the radial lines.

For the half-wave antenna it is sufficient to consider the pattern in but two of the principal planes—the plane perpendicular to the antenna and bisecting it and any plane containing the antenna. These are called, respectively, the magnetic plane and the electric plane, because the former contains magnetic-flux lines and the latter contains electric-flux lines. These names are very convenient in dealing with single half-wave antennas but are less convenient in dealing with other antenna systems. When used with other systems, the term *electric plane* usually means a

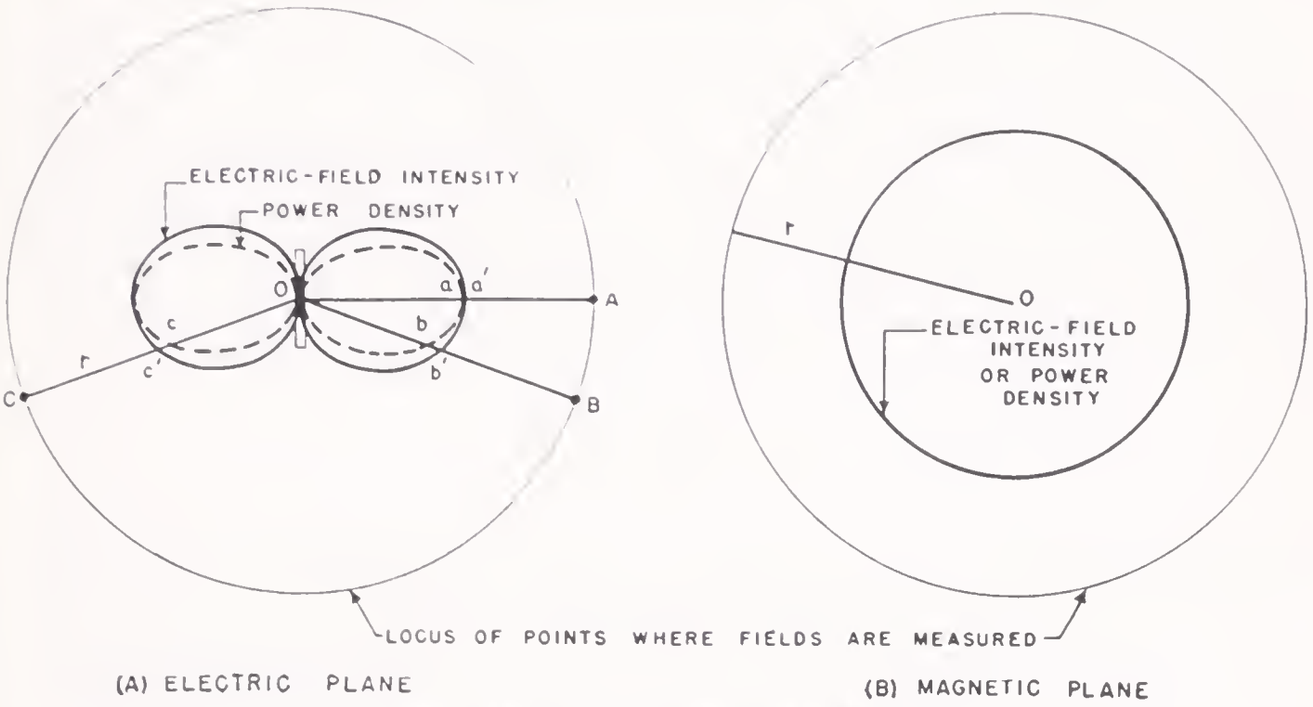


FIG. 14. Patterns for a half-wave antenna.

plane containing electric-flux lines and the line of principal radiation. Likewise, the term *magnetic plane* usually means the plane containing magnetic-flux lines and the line of principal radiation.

Figure 14*A* shows the electric-field and power-density patterns for a half-wave antenna in the electric plane, and Fig. 14*B* shows the patterns in the magnetic plane. The plots are polar plots, in which the length ob is proportional to $\mathcal{S}_{rm}$ at point B and the length ob' is proportional to $\mathcal{E}_{\theta m}$ at point B, and so on, for other points such as A and C. The scales have been so chosen that the electric-field and power-density patterns coincide for point A, which is a point where both $\mathcal{E}_{\theta m}$ and $\mathcal{S}_{rm}$ are maximum. This procedure makes it possible to observe that $\mathcal{S}_{rm}$ falls away from its maximum value more rapidly with change in angle than does $\mathcal{E}_{\theta m}$. In fact the $\mathcal{S}_{rm}$ pattern may be obtained by squaring values read from the $\mathcal{E}_{\theta m}$ pattern when the scales are such that oa and oa' are unity. An important consequence of this is that $\mathcal{S}_{rm}$ reaches half its maximum

value where $\mathcal{E}_{\theta m}$ is 0.707 of its maximum value. Points on the pattern which satisfy this condition are called *half-power points*. The angle between the half-power points is defined as the *beam angle*. The beam angle of a half-wave antenna is 78.2 deg in the electric plane.

Because the patterns for all planes containing the half-wave antenna are identical, it is possible to visualize a three-dimensional pattern which is obtained by revolving the $\mathcal{E}_{\theta m}$ curve of Fig. 14*A* around the antenna as an axis, thereby generating a doughnut-shaped surface which has no hole. The radial distance from the center to any point on this surface is proportional to $\mathcal{E}_{\theta m}$ at a specified distance r measured in the direction of plotting. A three-dimensional power-density pattern can also be generated, of course.

The above statements regarding patterns have presumed that the radial distance r is constant and that $\mathcal{E}_{\theta m}$ and $\mathcal{S}_{rm}$ are varying with the angles. Another interpretation can, however, be made. If contour lines are drawn in a plane by joining the points at which $\mathcal{E}_{\theta m}$ has certain specified values, their shape is identical in every respect with the electric-field pattern discussed above. This can be seen by reversing the positions of r and $\mathcal{E}_{\theta m}$ in the expression previously given for $\mathcal{E}_{\theta m}$; thus

$$r = \frac{60 I_m}{\mathcal{E}_{\theta m}} F_0(\theta)$$

When $\mathcal{E}_{\theta m}$ is fixed in this expression, r varies in proportion to $F_0(\theta)$, which is the way that $\mathcal{E}_{\theta m}$ varies when r is fixed. It should be observed, however, that power-density patterns are not identical in shape with contours of constant power density, since power density decreases with the square of the distance. Contours of fixed power density can be drawn, but their shape is the same as that of the contours of fixed $\mathcal{E}_{\theta m}$, because fixing $\mathcal{S}_{rm}$ also fixes $\mathcal{E}_{\theta m}$.

Another possibility in plotting patterns is to express the values of $\mathcal{E}_{\theta m}$ or $\mathcal{S}_{rm}$ as a fraction or as a percentage of the largest value of the quantity being plotted. Then the pattern has a maximum magnitude of unity or 100 per cent, and the result is not restricted to any particular radial distance. The pattern shows the angular variational features only, and these are the same for any radial distance. It should be noted that the power-density patterns are a measure of the power radiated per unit solid angle in the various directions.

The directional pattern of a receiving antenna is a plot of the antenna output as a function of the direction of approach of a constant-intensity field. If output *voltage* is plotted, patterns analogous to the field-intensity patterns of transmitting antennas are obtained; if output *power* is plotted, patterns analogous to power-density patterns are obtained.

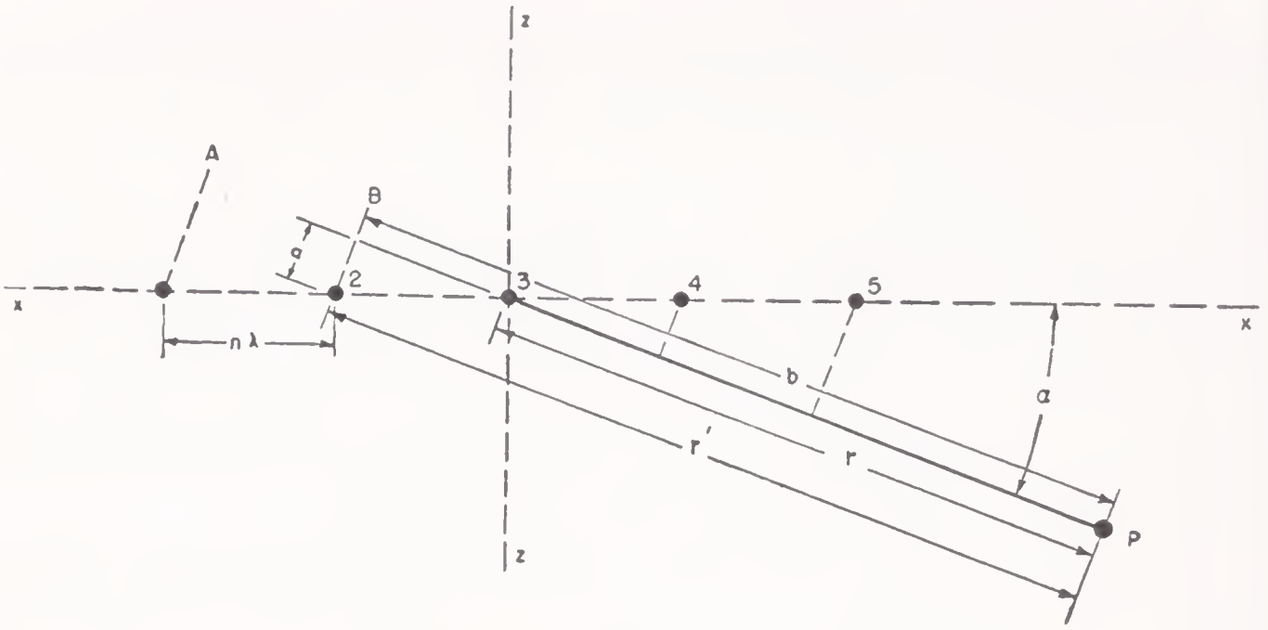
Absolute patterns showing voltage or power output for unit intensity of the incident field may be drawn, or if only directional information is required, per-unit patterns showing output as a fraction of maximum output may be used.

The over-all pattern of a radar antenna is the product of the transmitting-antenna pattern and the receiving-antenna pattern. Since in most radar units the same antenna is used for transmitting and for receiving, the over-all or round-trip pattern per-unit form is the square of the individual or one-way pattern. Since the square of the per-unit field pattern is the per-unit power-density pattern, the round-trip field pattern and the one-way power-density pattern are identical.

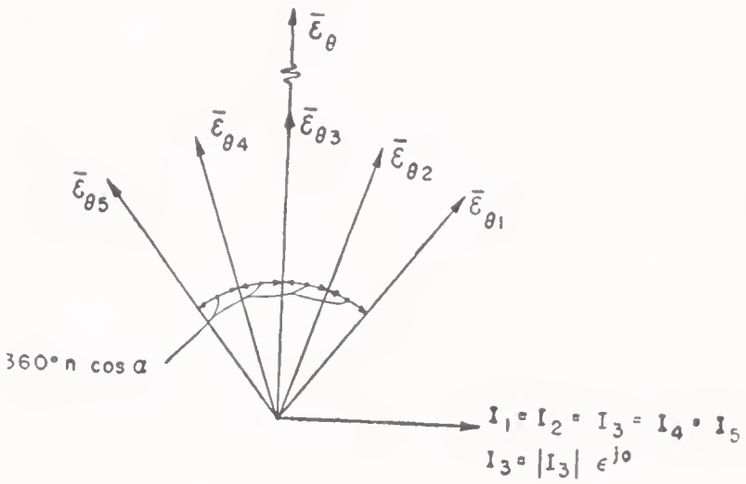
6. Radiation Pattern of an Array of Half-wave Antennas Fed in Phase. Thus far this chapter has dealt with only a single half-wave radiating element. Radar antennas operating in the ultrahigh-frequency region are usually composed of a number of half-wave elements arranged in symmetrical geometric manner. For example, several half-wave radiators may be placed parallel to one another with equal spacing between their centers, or a curtain type of array may be built up by arranging a number of half-wave elements parallel to one another in rows and columns, with equal spacing between adjacent rows and columns. Sometimes two curtains are employed, one behind the other. Through proper spacing of the elements and phasing of the currents in the elements, a directional radiation pattern having a beam angle small compared with that of a single-half wave element is obtained. A method of determining the radiation pattern of an array of symmetrically arranged half-wave antennas is presented in this and the next two articles.

Assume that several half-wave antennas, all lying in one plane, are arranged with their centers along a straight line, as shown in Fig. 15*A*. Assume also that the spacing between centers of adjacent antennas is $n\lambda$ and that the currents in all the antennas are in phase, have equal magnitude, and have the same reference direction. The instantaneous electric-field intensity in the θ direction contributed at point P by each half-wave antenna may be calculated from Eq. 1, p. 908, provided the frequency and radial distance between each antenna and point P are known. Because the radial distances are slightly different for each antenna, the contributions of field at point P made by the antennas are delayed by different amounts during propagation. The resultant field in the θ direction at P is therefore the sum of several component fields which are out of phase with one another.

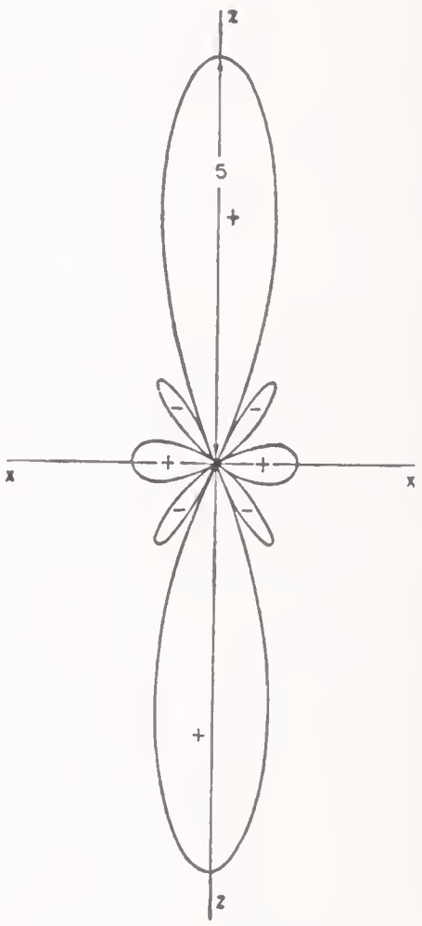
The resultant field at P may be determined by constructing a complex-number diagram, as shown in Fig. 15*B*. Since all the antenna currents are of equal magnitude and are in phase, they may be represented by the complex number $I = (I_m/\sqrt{2})e^{j0}$. In Fig. 15*B*, I is taken as the refer-



(A) ARRANGEMENT OF ELEMENTS



(B) COMPLEX-NUMBER DIAGRAM



(C) RADIATION PATTERN FOR n = 1/2

FIG. 15. Antenna array formed by spacing half-wave elements at equal intervals along a line through their centers. The currents in all elements have equal amplitudes and are in phase.

ence wave. The complex number $\bar{\mathcal{E}}_{\theta 3}$, representing the field laid down at P by the current in antenna 3, has a magnitude, from Eq. 4, of

$$|\bar{\mathcal{E}}_{\theta 3}| = \frac{\mathcal{E}_{\theta m 3}}{\sqrt{2}} = \frac{60 I_m}{\sqrt{2}\, r} F_0(\theta)$$

According to Eq. 1, $\bar{\mathcal{E}}_{\theta 3}$ lags I by the angle $360° (r/\lambda) - 90°$. Hence

$$\bar{\mathcal{E}}_{\theta 3} = \frac{\mathcal{E}_{\theta m 3}}{\sqrt{2}} \epsilon^{-j\left(360° \frac{r}{\lambda} - 90°\right)} \tag{7}$$

If r/λ is assumed to be an integer, $\bar{\mathcal{E}}_{\theta 3}$ is delayed an integral number of revolutions less 90 deg from I and therefore appears 90 deg ahead of I on the diagram of Fig. 15*B*.

The radial distance r' between antenna 2 and point P is slightly greater than r. It may be expected, therefore, that the field laid down at P by the current in antenna 2 lags slightly the field laid down at the same point by the current in antenna 3. Distance r' is the hypotenuse of a right triangle of which distance $b = r + n\lambda \cos \alpha$ is the base, and the normal distance a from antenna 2 to the base is the altitude; that is,

$$r' = \sqrt{b^2 + a^2} = b + \frac{a^2}{2b} - \frac{a^4}{8b^3} + \cdots \tag{8}$$

If point P is remote from the array, as it must be if only radiation fields are of importance, distance b is large compared with a, and r' may be taken with negligible error as

$$r' = b = r + n\lambda \cos \alpha \tag{9}$$

The error introduced by such an approximation may be determined by noting that, in an alternating series such as that of Eq. 8, the error introduced by neglecting all terms beyond any one in the expansion is not greater than the first term neglected. Thus the difference between r' and b is no greater than $a^2/2b$. In order for this difference to be less than $\lambda/360$, so that error in the measurement of phase of $\bar{\mathcal{E}}_{\theta 2}$ may be less than 1 deg, b must be greater than $180a^2/\lambda$. Usually calculations are made at such distances from antenna arrays that this condition is easily satisfied. Thus it may be assumed that the radial distance from antenna 2 to point P is given by Eq. 9, and $\bar{\mathcal{E}}_{\theta 2}$ is therefore

$$\bar{\mathcal{E}}_{\theta 2} = |\bar{\mathcal{E}}_{\theta 2}| \epsilon^{-j\left[\frac{360°(r + n\lambda \cos \alpha)}{\lambda} - 90°\right]}$$

Because $n\lambda \cos \alpha$ is usually very small compared with r, the quantity $n\lambda$ may be neglected in a calculation of $|\bar{\mathcal{E}}_{\theta 2}|$, and hence

$$|\bar{\mathcal{E}}_{\theta 2}| = |\bar{\mathcal{E}}_{\theta 3}| = \frac{\mathcal{E}_{\theta m 3}}{\sqrt{2}}$$

Thus the complex number $\bar{\mathcal{E}}_{\theta 2}$, representing the electric field laid down at P by the current in antenna 2, is

$$\bar{\mathcal{E}}_{\theta 2} = \frac{\mathcal{E}_{\theta m 3}}{\sqrt{2}} \epsilon^{-j\left(360°\frac{r}{\lambda}+360°n\cos\alpha-90°\right)} \tag{10}$$

Comparison of Eqs. 7 and 10 shows that $\bar{\mathcal{E}}_{\theta 2}$ lags $\bar{\mathcal{E}}_{\theta 3}$ by the angle $360°n \cos \alpha$, as indicated in Fig. 15*B*.

The radial distance between antenna 1 and point P, in accordance with the approximation made above, is $r + 2n\lambda \cos \alpha$. If $2n\lambda \cos \alpha$ is neglected in the calculation of magnitude, $\bar{\mathcal{E}}_{\theta 1}$, the complex number standing for the field laid down at P by antenna 1, is

$$\bar{\mathcal{E}}_{\theta 1} = \frac{\mathcal{E}_{\theta m 3}}{\sqrt{2}} \epsilon^{-j\left(360°\frac{r}{\lambda}+360°2n\cos\alpha-90°\right)}$$

Thus $\bar{\mathcal{E}}_{\theta 1}$ lags $\bar{\mathcal{E}}_{\theta 3}$ by the angle $2(360°n \cos \alpha)$ and $\bar{\mathcal{E}}_{\theta 2}$ by the angle $360°n \cos \alpha$.

In a similar manner it may be shown that $\bar{\mathcal{E}}_{\theta 4}$ and $\bar{\mathcal{E}}_{\theta 5}$ have magnitudes equal to the magnitude of $\bar{\mathcal{E}}_{\theta 3}$, that $\bar{\mathcal{E}}_{\theta 4}$ leads $\bar{\mathcal{E}}_{\theta 3}$ by the angle $360°n \cos \alpha$, and that $\bar{\mathcal{E}}_{\theta 4}$ leads $\bar{\mathcal{E}}_{\theta 5}$ also by an equal angle. Thus the complex-number diagram for the electric field in the θ direction takes the form shown in Fig. 15*B*. Observe that all complex numbers have equal magnitudes and that all are displaced in phase by the angle $360°n \cos \alpha$ because of the difference in radial distances between point P and the component elements of the array.

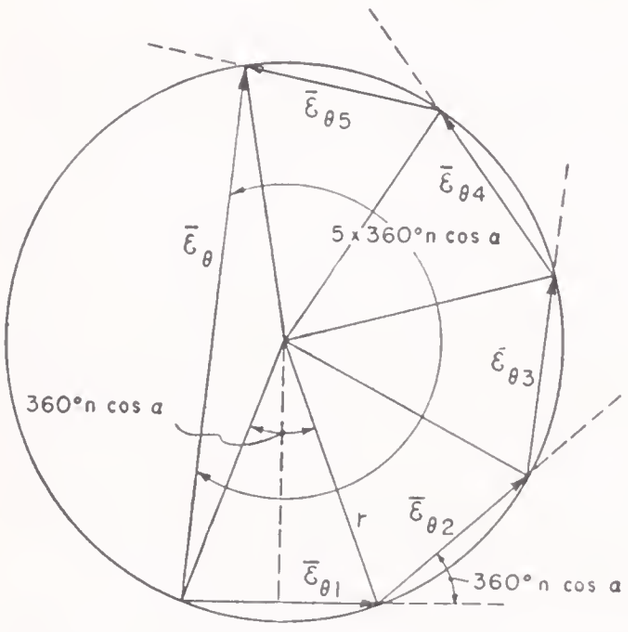


FIG. 16. Graphical addition of complex numbers represented in Fig. 15*B*.

The resultant field at P is the complex-number sum of all the component fields contributed at P by the various elements of the array. This summation is performed graphically in Fig. 16. Each complex-number magnitude $\mathcal{E}_{\theta m 3}/\sqrt{2}$ constitutes a chord of a circle, and $\mathcal{E}_{\theta m 3}/\sqrt{2}$ is related to the circle radius r and the angle subtended by the chord in the following manner:

$$\frac{\mathcal{E}_{\theta m 3}}{\sqrt{2}} = 2r \sin\left(\frac{360°n \cos \alpha}{2}\right) \tag{11}$$

The magnitude of the resultant complex number $\bar{\mathcal{E}}_\theta$ is also a chord of the same circle and has a magnitude

$$\frac{\mathcal{E}_{\theta m}}{\sqrt{2}} = 2r \sin \frac{5(360°n \cos \alpha)}{2} \tag{12}$$

Thus, from Eqs. 11 and 12,

$$\mathcal{E}_{\theta m} = \mathcal{E}_{\theta m 3} \frac{\sin 5(180°n \cos \alpha)}{\sin 180°n \cos \alpha} \tag{13}$$

that is, *the peak value of the resultant field at point P is equal to the peak value of the field contributed by one antenna multiplied by a sine-over-sine factor.*

The phase of the resultant field may be determined directly from Fig. 15*B*. If the separation angles between adjacent complex numbers is not larger than 360°/5, $\bar{\mathcal{E}}_\theta$ is in phase with $\bar{\mathcal{E}}_{\theta 3}$. If the separation angle exceeds 360°/5 and does not exceed 2 × 360°/5, $\bar{\mathcal{E}}_\theta$ is opposite in phase to $\bar{\mathcal{E}}_{\theta 3}$. Values of α which made the resultant field opposite to $\bar{\mathcal{E}}_{\theta 3}$ also make the numerical value of the sine-over-sine factor in Eq. 13 negative. Thus it may be stated that the phase of the resultant field is the same as the phase of the contribution made by a single antenna located at the center of the group when the sine-over-sine factor is positive and opposite phase when the sine-over-sine factor is negative.

Equation 13 may be used to plot the radiation pattern from the five-element array of Fig. 15*A* in the *xz* plane. Values of α are assumed, and the corresponding ratios of $\mathcal{E}_{\theta m}/\mathcal{E}_{\theta m 3}$ are calculated. The result is shown in Fig. 15*C* for a half-wave spacing of the elements ($n = \frac{1}{2}$). Observe that the pattern contains a number of minor lobes in addition to two main lobes and that the beam angle of the main lobes is much smaller than the beam angle of a single half-wave element. The narrow beam angle and the side lobes are the result of the out-of-phase relation that exists among the component fields laid down at various points in the *xz* plane at fixed distances from the center of the array. The plus signs within the lobes signify that within the region covered by the lobes the resultant field is in phase with the field laid down by the element at the center of the array; the minus signs indicate a 180-deg phase relation between resultant field and the component field contributed by the center element of the array.

Equation 13, which relates the magnitude of the resultant field from a line of five elements to the magnitude of the field contributed by a single element, may be generalized to include a line of N elements equally spaced a distance $n\lambda$. For a line of N equally spaced elements fed in phase,

the peak value of the resultant electric-field intensity $\mathcal{E}_{\theta m}$ at a point remote from the array becomes

$$\mathcal{E}_{\theta m} = \mathcal{E}_{\theta m1} \frac{\sin N(180°n \cos \alpha)}{\sin (180°n \cos \alpha)} \tag{14}$$

where $\mathcal{E}_{\theta m1}$ is the peak value of the field in the θ direction contributed by any one element. The validity of this relation may be demonstrated by extending the graphical addition in Fig. 16 to include N equally spaced elements. For convenience, the sine-over-sine factor in Eq. 14 may be expressed in the following manner:

$$\frac{\mathcal{E}_{\theta m}}{\mathcal{E}_{\theta m1}} = F^{Nn}(\alpha) = \frac{\sin N(180°n \cos \alpha)}{\sin (180°n \cos \alpha)} \tag{15}$$

where the factor $F^{Nn}(\alpha)$ signifies the sine-over-sine factor for a line of half-wave antennas spaced n wavelengths apart for various values of the angle α, which is the angle included between a radial line from the center of the group and a line through their centers. The F factors given by Eq. 15 have particular importance in that they characterize the radiation patterns of multiple-element arrays fed in phase. They specify the manner in which the radiation pattern of a single half-wave element is modified because of the presence of additional elements.

Properties of the F factors are most easily explained by considering Fig. 17, which shows a selection of them plotted in polar coordinates. The factors are not plotted to the same scale, but adjustment in scale is easily comprehended by observing that the lobes at right angles to the line where α is zero have a peak magnitude of N. In this direction all radiations travel the same distance, so that they arrive in phase because they started in phase; hence the resultant must be N times the contribution from a single source. It is important to note that the factors are identical on either side of the line where α is zero. This means that angles can be measured from either side of the line of centers. Likewise it will be observed that the factors are symmetrical about a line perpendicular to the line where α is zero. The angle α can therefore be measured from either end of the line of centers. Because of this symmetry it is unnecessary to tabulate F factors over more than a 90-deg range of angles.

From Fig. 17 two trends may be observed. The first trend is that the main lobes, which occur where α is 90 or 270 deg, are sharpened either by increasing the number N of elements or by increasing the spacing n between elements. That is, the sharpness of these lobes is principally dependent on the total length of the line of radiators, which is $n(N - 1)$. This fact can be seen by comparing the sharpness of the main lobes of $F^{2\frac{1}{2}}$ and $F^{4\frac{1}{4}}$; $F^{2\frac{3}{4}}$, $F^{4\frac{1}{2}}$, and $F^{8\frac{1}{4}}$; $F^{4\frac{3}{4}}$ and $F^{8\frac{1}{2}}$. The second trend is that the number of minor lobes increases as the sharpness of the main lobes

increases. However, the magnitude of the largest minor lobe, if expressed as a percentage of the main lobe, is roughly independent of the sharpness of the main lobe. This, however, is true only for spacings somewhat below one wavelength. When the spacing is made equal to one wavelength, the value of the factor is as great at $\alpha = 0$ as at $\alpha = 90$ deg. Spacings are kept below three-fourths wavelength to avoid this end-fire effect. The end-fire effect in the lower row of diagrams in Fig. 17 occurs

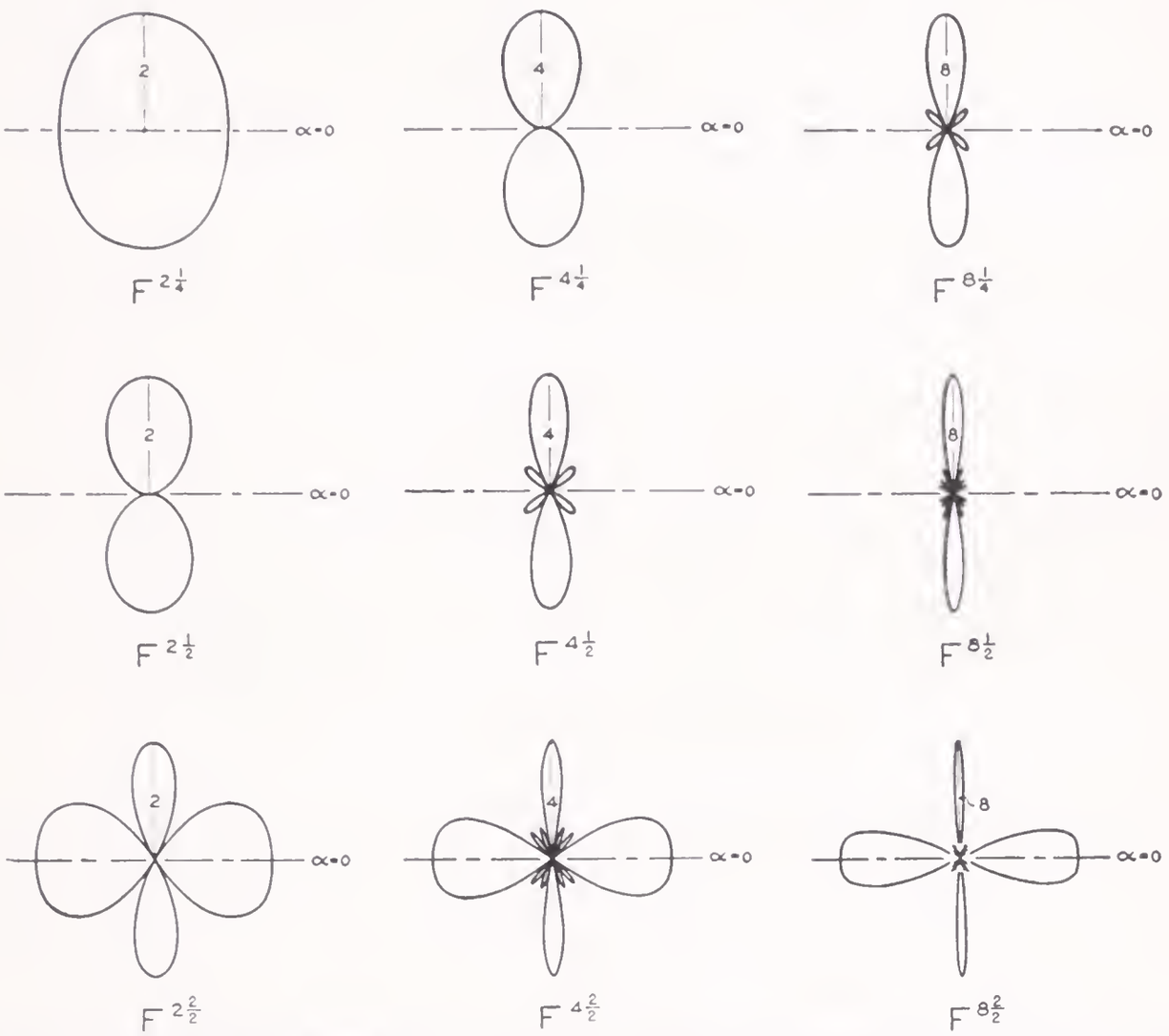
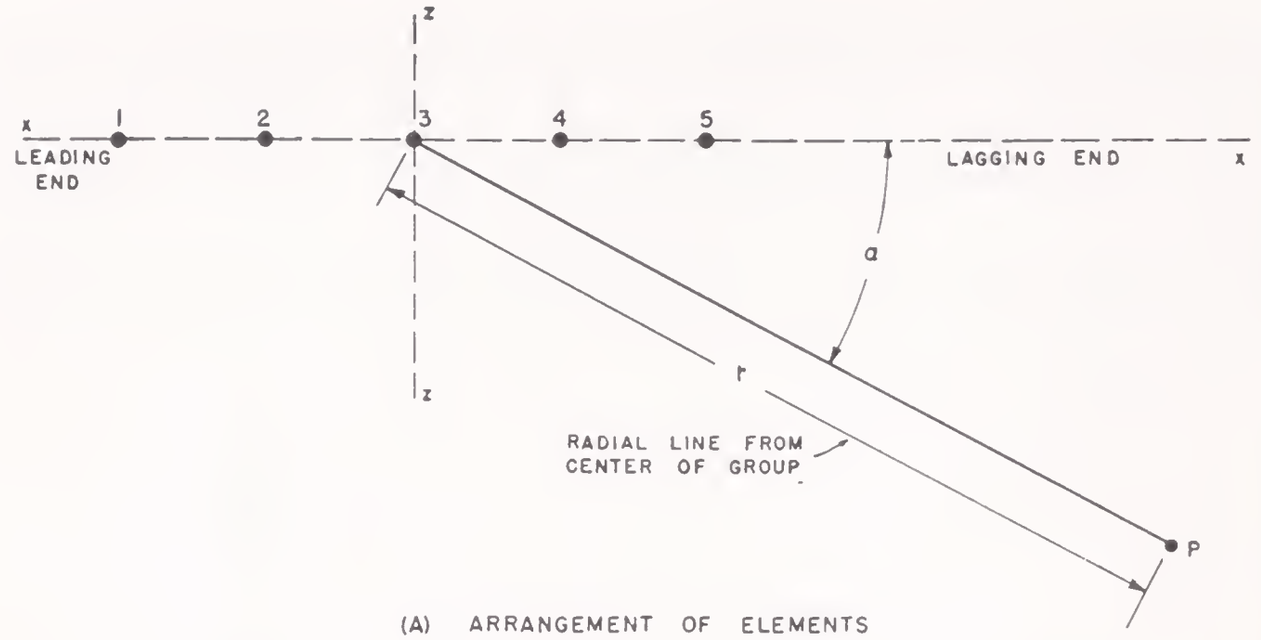


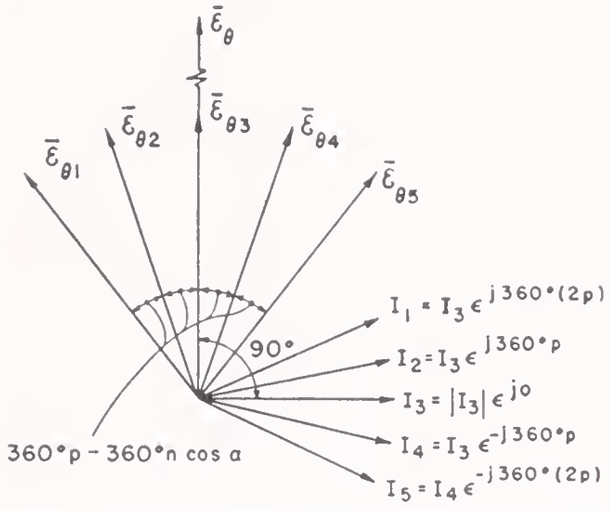
FIG. 17. Polar plots of F factors for two, four, or eight sources excited in phase and spaced $\lambda/4$, $\lambda/2$, or λ.

because one cycle of radiation from a given antenna, when delayed a cycle due to one wavelength of extra distance traveled along the line of the array, arrives in phase with the next cycle of radiation from the next antenna, and so on.

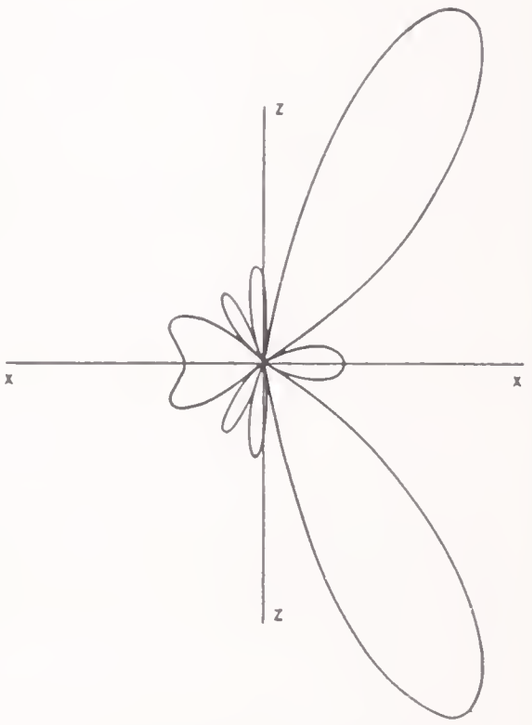
7. Radiation Pattern of an Array of Half-wave Antennas Fed out of Phase. Consider now the same group of antennas shown in Fig. 15, and assume that the currents in all elements are of equal magnitude but that the current in antenna 1 leads the current in antenna 2 by p cycles, the current in antenna 2 leads the current in antenna 3 by p cycles, and so on. The angle of phase difference between currents in adjacent elements is



(A) ARRANGEMENT OF ELEMENTS



(B) COMPLEX-NUMBER DIAGRAM



(C) XZ-PLANE RADIATION PATTERN FOR n=1/4 AND n=1/2

FIG. 18. Complex-number diagram and radiation pattern of antenna array of Fig. 16 for the conditions $n = \frac{1}{2}$ and $p = \frac{1}{4}$.

therefore $360p$ deg. The arrangement of elements is shown in Fig. 18*A*, and the complex-number diagram representing the component electric fields in the θ direction laid down by the array at point P in the xz plane is given in Fig. 18*B*. This diagram is constructed in accordance with the principles presented in the preceding article, the angle of phase difference between currents in adjacent elements being taken into account.

In Fig. 18*B* the current in antenna 3 is taken as the reference wave, and distance r between P and the center of the array is assumed to be an integral number of wavelengths. Thus $\bar{\mathcal{E}}_{\theta 3}$ lags I_3 by a whole number of

revolutions less 90 deg and appears in Fig. 18*B* 90 deg ahead of I_3. Similarly, $\bar{\mathcal{E}}_{\theta 2}$ lags I_2 by a whole number of revolutions less

$$90^\circ - 360^\circ n \cos \alpha$$

Because I_2 *leads* I_3 by the angle $360^\circ p$, $\bar{\mathcal{E}}_{\theta 2}$ and $\bar{\mathcal{E}}_{\theta 3}$ are displaced in phase by the angle $360^\circ p - 360^\circ n \cos \alpha$. It is assumed in Fig. 18*B* that angle $360^\circ p$ is large enough so that $\bar{\mathcal{E}}_{\theta 2}$ leads $\bar{\mathcal{E}}_{\theta 3}$, as shown in the diagram. By similar reasoning it may be shown that the remaining complex numbers $\bar{\mathcal{E}}_{\theta 1}$, $\bar{\mathcal{E}}_{\theta 4}$, and $\bar{\mathcal{E}}_{\theta 5}$ are positioned as shown in Fig. 18*B*, the angle of phase difference between adjacent complex numbers being

$$360^\circ p - 360^\circ n \cos \alpha$$

When the component fields at P are summed, the peak value of the resultant field is given by Eq. 13 with the angle $180^\circ n \cos \alpha$ replaced by $180^\circ p - 180^\circ n \cos \alpha$; that is,

$$\mathcal{E}_{\theta m} = \mathcal{E}_{\theta m 3} \frac{\sin 5(180^\circ p - 180^\circ n \cos \alpha)}{\sin (180^\circ p - 180^\circ n \cos \alpha)} \tag{16}$$

The radiation pattern in the xz plane for the five-element array of Fig. 18*A* is drawn in Fig. 18*C* for a one-half-wavelength spacing between elements and a phasing of one-quarter cycle between adjacent elements. Thus the pattern is a plot of

$$\mathcal{E}_{\theta m} = \mathcal{E}_{\theta m 3} \frac{\sin 5(45^\circ - 90^\circ \cos \alpha)}{\sin (45^\circ - 90^\circ \cos \alpha)}$$

Observe that the xz-plane pattern is symmetrical about the line of centers, as in Fig. 16*C*, but is unsymmetrical about a line perpendicular to the line of centers. Because of the dissymmetry, angle α must be measured from the lagging end of the line of centers, and values must be computed through a 180-deg range of α. Because of the symmetry about the line of centers, α may be measured in either a clockwise or counterclockwise direction.

Equation 16 can be generalized to include the case of a line of N antennas spaced a distance $n\lambda$ apart with a phasing of p cycles between adjacent elements. The peak value of the resultant field intensity is then

$$\mathcal{E}_{\theta m} = \mathcal{E}_{\theta m 1} \frac{\sin N(180^\circ p - 180^\circ n \cos \alpha)}{\sin (180^\circ p - 180^\circ n \cos \alpha)} \tag{17}$$

In Eq. 17, $\mathcal{E}_{\theta m 1}$ is the peak value of the electric field laid down at a point by any one of the antennas, N is the number of half-wave antennas in line, n is the spacing in wavelengths, p is the phasing between adjacent currents in cycles, and α is the angle between a radial line from the center of the group and a line through the centers measured from the

lagging end. Often the sine-over-sine factor in Eq. 17 is expressed in terms of a G factor in the following manner:

$$\frac{\mathcal{E}_{\theta m}}{\mathcal{E}_{\theta m1}} = G_p{}^{Nn}(\alpha) = \frac{\sin N(180°p - 180°n \cos \alpha)}{\sin (180°p - 180°n \cos \alpha)} \tag{18}$$

The G factors characterize the radiation patterns of a line of half-wave elements fed out of phase and therefore perform the same function as the F factors (Art. 6), which specify the radiation patterns when the elements are fed in phase. The G factors specify the manner in which the radiation pattern of a single half-wave element is modified because of the presence of additional elements and because of a difference in phase between currents in adjacent elements.

The G factors apply to any number of half-wave elements, whether even or odd. If the number of antennas is even, so that an antenna is not actually located at the center of the group, the phase of the resultant field is determined by considering a fictitious radiator to be placed at the center, its current being midway in phase between the currents on either side. A positive-valued G factor indicates that the resultant field is in phase with the field laid down by the current in the fictitious radiator; a negative G factor indicates opposite phase.

The effects on the G factors of changing the spacing of the antenna elements and of changing their phasing are illustrated by the polar plots of Fig. 19. These diagrams apply to two half-wave radiating elements spaced $\lambda/4$, $\lambda/2$, λ, and $3\lambda/2$ apart and fed out of phase by an amount 0, $\frac{1}{6}$, $\frac{2}{6}$, $\frac{3}{6}$, $\frac{4}{6}$, $\frac{5}{6}$, and 1 cycle. The effect of changing the spacing can best be seen by observing diagrams in the left-hand column. Here p is zero, so that the factors are really F factors. When $n = 0$, the plot is a circle of radius 2. This plot is not shown. When n is increased to $\frac{1}{4}$, the plot has a magnitude of 2 at 90 and 270 deg and a magnitude of 1.41 at 0 and 180 deg. The effect of increasing n, then, is to narrow the figure at the line of centers. Further increase of n toward $\frac{1}{2}$ is accompanied by an increased amount of narrowing until at $n = \frac{1}{2}$ the figure is divided. If n is still further increased, small lobes appear at $\alpha = 0$ and $\alpha = 180$ deg. These become full grown when $n = 1$ and then begin to divide as n goes beyond 1. Complete division is achieved when $n = \frac{3}{2}$. Every additional wavelength increase in spacing is accompanied by the growth of lobes during the first half of the increase and by the division of these during the latter half. This means that at spacings of $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$ wavelengths there will be 1, 3, 5, and so on, complete lobes in the range from $\alpha = 0$ to $\alpha = 180$ deg.

If, now, n is fixed at some value, say $\frac{3}{2}$, and p is varied, the factors behave in a simple way. When p is zero, the pattern has three distinct lobes on each side of the line of centers. When p is increased, all these

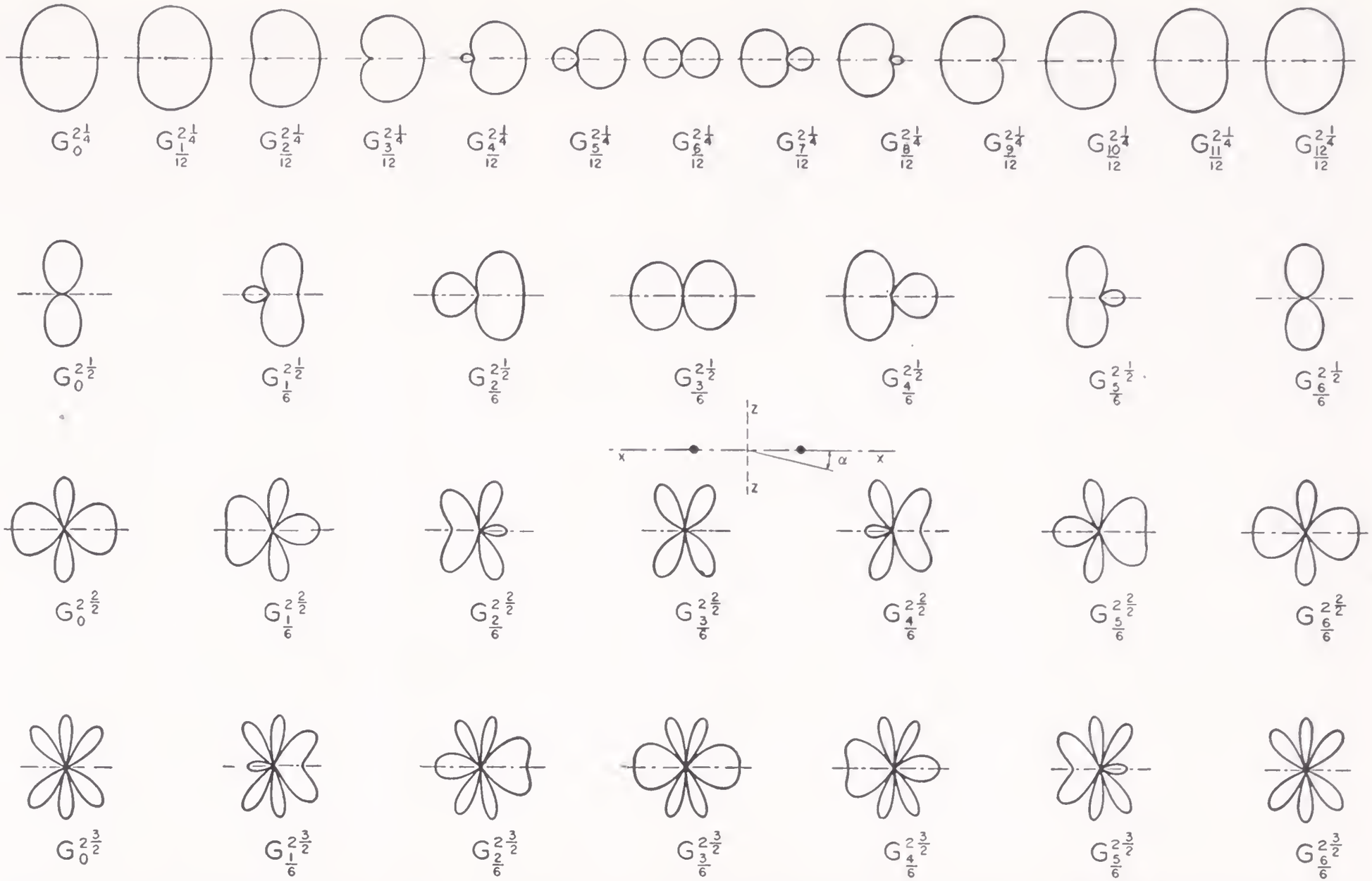


FIG. 19. Polar plots of G factors for two sources spaced $\lambda/4$, $\lambda/2$, λ, and $3\lambda/2$ and fed out of phase by 0, $\frac{1}{6}$, $\frac{2}{6}$, $\frac{3}{6}$, $\frac{4}{6}$, $\frac{5}{6}$, and 1 cycle.

lobes begin to migrate toward the lagging end of the line of centers ($\alpha = 0$), and a small one begins to appear at $\alpha = 180$ deg. The migration of the two lobes at the right causes them to coalesce, as indicated. The figures for $p = \frac{1}{6}$, $p = \frac{2}{6}$, and $p = \frac{3}{6}$ cycle show this effect in varying degrees. But at $p = \frac{3}{6}$ the small lobe is full grown and the coalescing lobes appear as a single lobe. Further increase of p beyond $\frac{3}{6}$ is accompanied by a gradual division of the full-grown lobe and the vanishing of the coalesced lobes—a process which is complete when $p = \frac{6}{6}$. The plots when $p = 0$ and $p = \frac{6}{6}$ are identical in shape, but the identity of the lobes has changed. Every lobe has advanced to

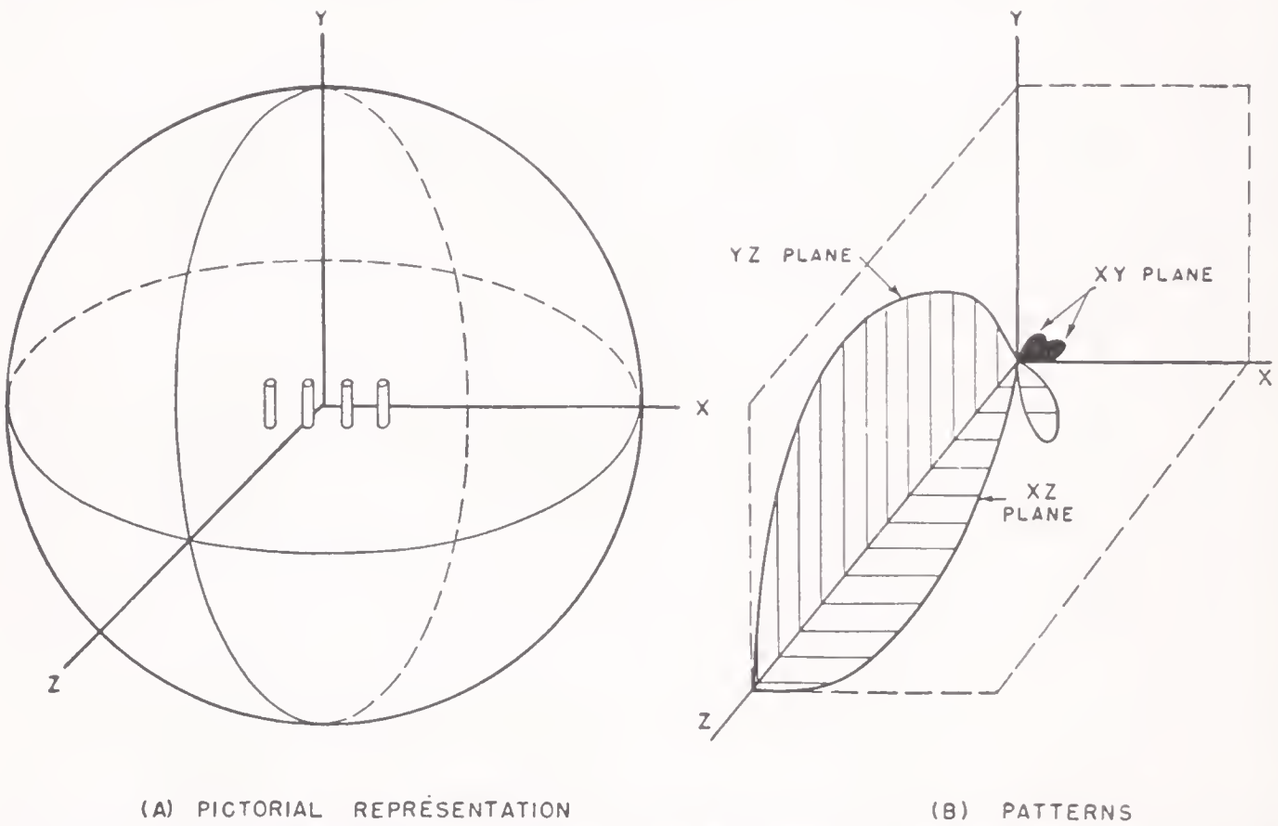


FIG. 20. Radiation patterns, in the three principal planes, of four half-wave antennas fed in phase.

the next position nearer the lagging end of the line of centers. This turning of the lobes by changing p is an important property of the G factors.

8. Examples of the Use of Factors in Obtaining Antenna Patterns. Several examples illustrating the use of F factors for obtaining antenna patterns are given in this article.

Patterns from a Row of Half-wave Antennas. Assume that four half-wave antennas are arranged in the xy plane with centers uniformly spaced at intervals of $\lambda/2$ along the x axis, as indicated in Fig. 20*A*. Each carries current of magnitude I_m with a reference direction along the positive y axis, and all currents are in phase. Patterns are required in the three principal planes—the xz plane, the yz plane, and the xy plane. For such an array the four necessary conditions for using factors are satisfied. Contributions from individual antennas are of equal magni-

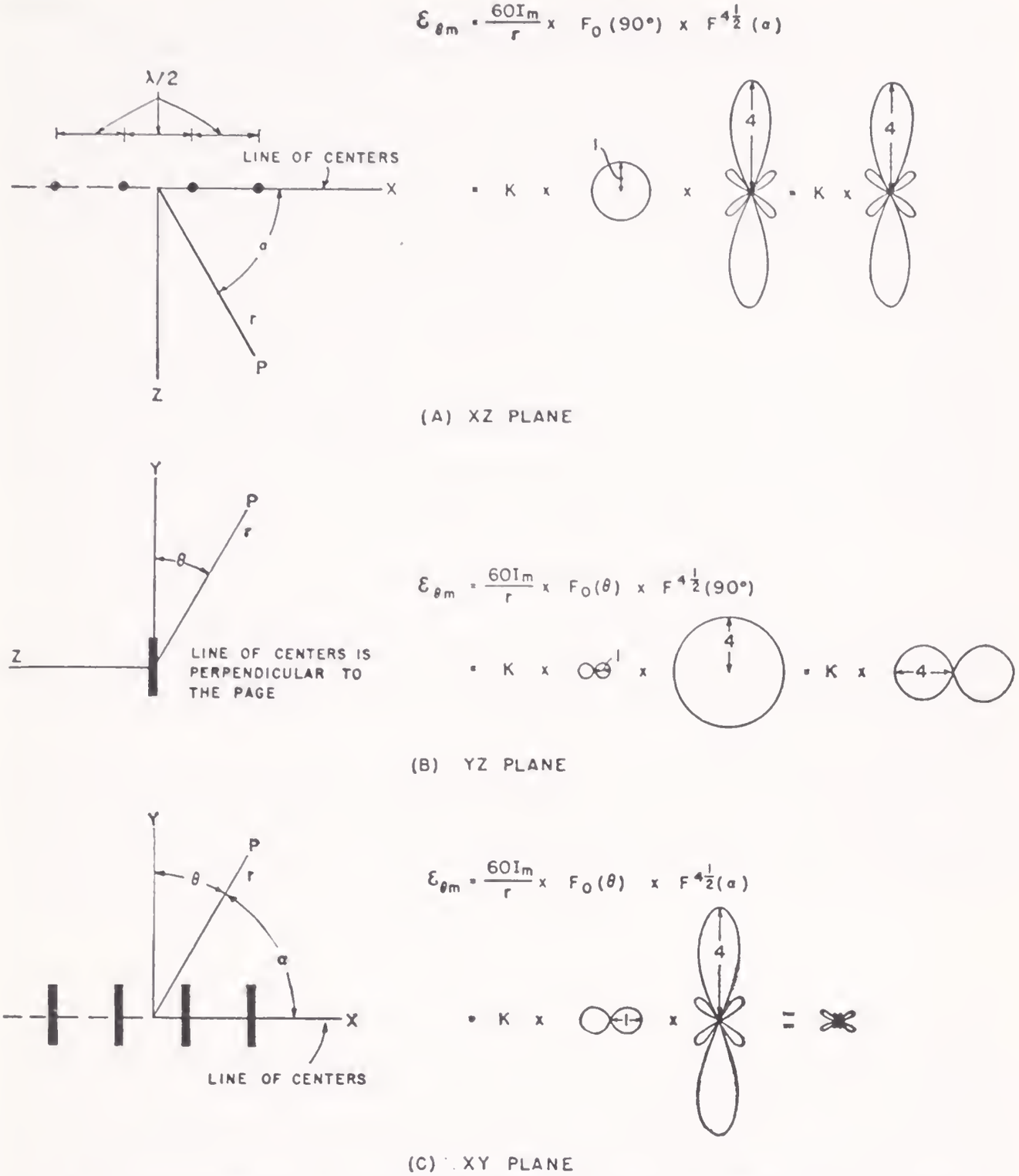

$\mathcal{E}_{\theta m} = \frac{60I_m}{r} \times F_0(90°) \times F^{4\frac{1}{2}}(\alpha)$
λ/2
LINE OF CENTERS
X
Z
P
r
α
= K x 1 x 4 = K x 4
(A) XZ PLANE
$\mathcal{E}_{\theta m} = \frac{60I_m}{r} \times F_0(\theta) \times F^{4\frac{1}{2}}(90°)$
Y
θ
LINE OF CENTERS IS PERPENDICULAR TO THE PAGE
= K x 1 x 4 = K x 4
(B) YZ PLANE
$\mathcal{E}_{\theta m} = \frac{60I_m}{r} \times F_0(\theta) \times F^{4\frac{1}{2}}(\alpha)$
= K x 1 x 4 =
(C) XY PLANE

FIG. 21. Space diagrams, pattern equations, and sketches of patterns for antennas of Fig. 20.

tude, the θ directions are identical, the spacing is uniform along a straight line, and the phasing is uniform.

The pattern in the xz plane is obtained most easily by consideration of Fig. 21A. In this plane the contribution of electric field from any one half-wave antenna at point P is $(60I_m/r)F_0(90°)$, and the contribution from all four is $F^{4\frac{1}{2}}(\alpha)$ times as large. Thus the resultant field at point P is given by the pattern equation

$$\mathcal{E}_{\theta m} = \frac{60I_m}{r} F_0(90°)F^{4\frac{1}{2}}(\alpha) = \text{constant} \times F^{4\frac{1}{2}}(\alpha)$$

The pattern therefore has the shape of the $F^{4\frac{1}{2}}$ factor, as indicated in Fig. 21*A*. Data for plotting this factor are obtained through the use of Eq. 15. Observe that each antenna by itself shows no directivity in the xz plane but that the four together show a very considerable amount of directivity. All the directivity is the result of the interference of the individual radiations; that is, the phase angles of the component fields at some points are proper for the effective addition of the fields, whereas at other points either partial or complete cancellation results.

The pattern in the yz plane can be obtained from Fig. 21*B*. Here the line of centers is perpendicular to the page and therfore makes a 90-deg angle with a radial line to point P, which necessarily lies in the plane of the page. The contribution from a single antenna is given by $(60I_m/r)F_0(\theta)$, and the contribution from the group is $F^{4\frac{1}{2}}(90°)$ times as large. The resultant field at point P is therefore

$$\mathcal{E}_{\theta m} = \frac{60I_m}{r} F_0(\theta)F^{4\frac{1}{2}}(90°) = \text{constant} \times F_0(\theta)$$

Thus the directivity comes entirely from the inherent directivity of the individual half-wave antennas, and the yz-plane pattern has the form indicated in Fig. 21*B*.

Finally, the pattern in the xy plane can be found by considering Fig. 21*C*. The contribution from a single antenna is $(60I_m/r)F_0(\theta)$, and the contribution from the group is $F^{4\frac{1}{2}}(\alpha)$ times as large. The resultant field is, accordingly,

$$\mathcal{E}_{\theta m} = \frac{60I_m}{r} F_0(\theta)F^{4\frac{1}{2}}(\alpha)$$

The pattern here is different from those in the other two principal planes in that both the directivity of the individual antennas and the interference between their contributions are involved. The factor $F^{4\frac{1}{2}}(\alpha)$ has its maximum value along the y axis, whereas the $F_0(\theta)$ function has its maximum value along the x axis. The product is zero along both the x and the y axes but has small values at intermediate positions.

An inspection of patterns for the three principal planes shows that the direction of principal radiation is along the z axis—a direction perpendicular to the plane containing the antennas. This is spoken of as *broadside* radiation. If the three patterns so far considered are placed together on a single diagram, an estimate of the three-dimensional pattern may be obtained. Figure 20*B* shows the three patterns as they appear in the first quadrant of each of the principal planes. From this figure, the three-dimensional pattern would be expected to consist of a main lobe directed along the z axis together with several minor lobes. The main lobe is much broader vertically than horizontally. A cross section taken parallel

to the xy plane would have an elliptical shape, the major axis being vertical or parallel to the y axis.

At points in the three principal planes where the $F^{4\frac{1}{2}}$ factor is positive, the phase of the resultant field is the same as it would be if the entire radiation had originated at a single point at the center of the group, which is the origin of coordinates in the present example. For points where the $F^{4\frac{1}{2}}$ factor is negative, the phase is opposite to the above. It should be noted that, in plotting patterns, the negative signs should be ignored, because patterns represent magnitude only and the negative sign refers to phase.

Simple Curtain of Half-wave Antennas. Another example of the use of multiple factors arises when several rows of antennas are arranged in a plane to form a vertical curtain, as indicated in Fig. 22. Here it is assumed that 12 identical half-wave antennas lie in the xy plane with spacings, measured from center to center, of $\lambda/2$ in both the x and the y direction. Each antenna carries a current of magnitude I_m with a reference direction along the positive x axis. All the currents are assumed to be in phase. In the present example the antennas lie horizontally in contrast to the preceding example where they are arranged vertically. The line of centers for the elements in a row is along the x direction, whereas the line of centers for the rows is along the y axis.

The equation of the pattern in the yz plane can be found by considering the curtain to consist of three identical rows. The contribution from each element in a row of four antennas is $(60I_m/r)F_0(\theta)$, and the contribution from all four elements in a row is

$$\frac{60I_m}{r} F_0(\theta)F^{4\frac{1}{2}}(\alpha_1)$$

where angle $\alpha_1 = \theta$ is the angle between a line through the centers of the elements in the row and a radial line from the center of the row and point P. These contributions from each of the three rows appear to come from the centers of the rows, which are points located along the y axis at $-\lambda/2$, 0, and $\lambda/2$. The angle between the y axis and a radial line from the center of the array to point P is α_2. Hence the rows may be added by multiplying the contribution from one row by a factor $F^{3\frac{1}{2}}(\alpha_2)$ to obtain

$$\mathcal{E}_{\theta m} = \frac{60I_m}{r} F_0(\theta)F^{4\frac{1}{2}}(\alpha_1)F^{3\frac{1}{2}}(\alpha_2)$$

A plot of the xy-plane pattern is given in Fig. 22*A*. The same result could have been obtained by considering the array as made up of four columns of three antennas each.

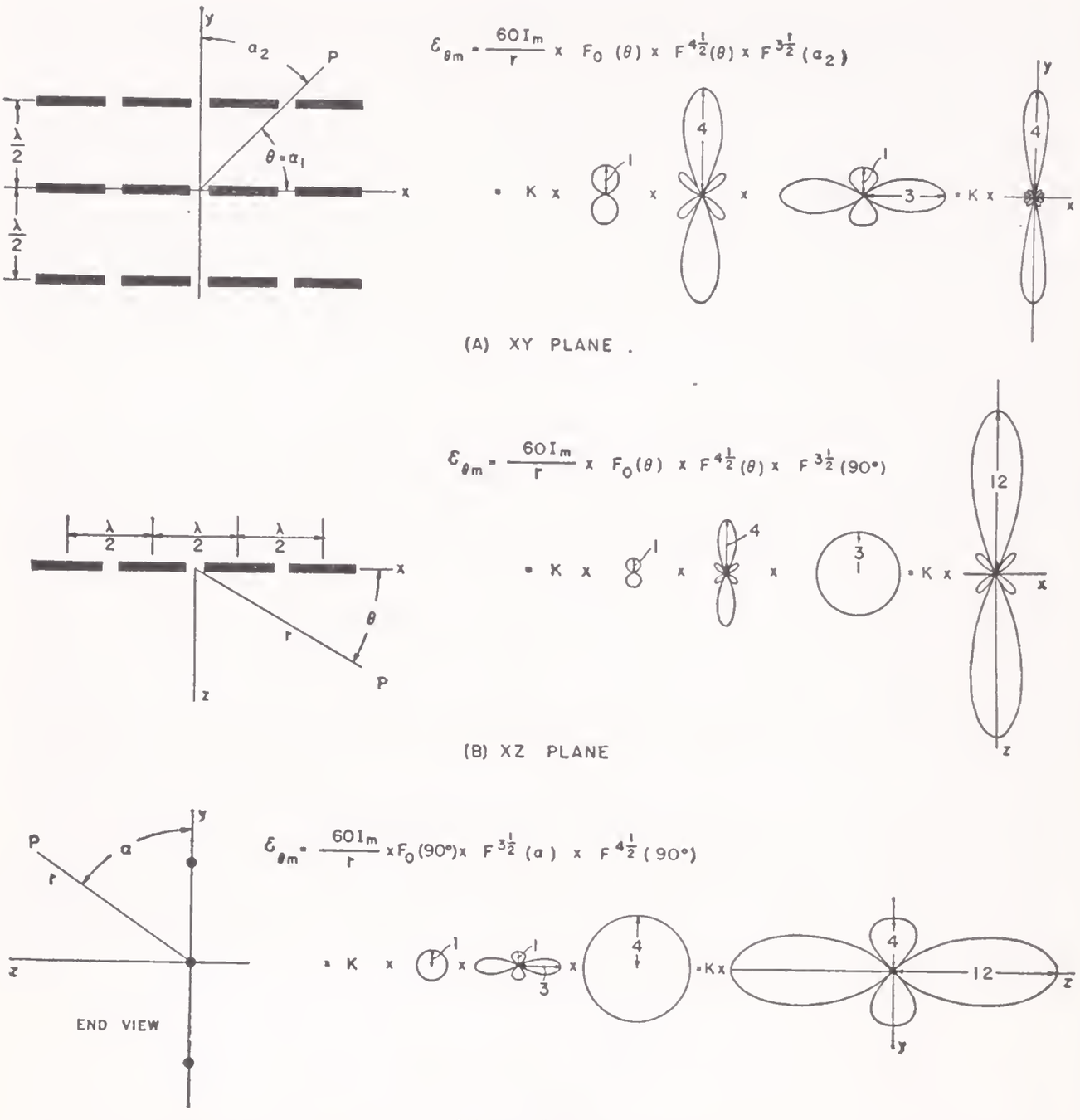

$\mathcal{E}_{\theta m} = \frac{60 I_m}{r} \times F_0(\theta) \times F^{4\frac{1}{2}}(\theta) \times F^{3\frac{1}{2}}(a_2)$
(A) XY PLANE
$\mathcal{E}_{\theta m} = \frac{60 I_m}{r} \times F_0(\theta) \times F^{4\frac{1}{2}}(\theta) \times F^{3\frac{1}{2}}(90°)$
(B) XZ PLANE
$\mathcal{E}_{\theta m} = \frac{60 I_m}{r} \times F_0(90°) \times F^{3\frac{1}{2}}(a) \times F^{4\frac{1}{2}}(90°)$
END VIEW
(C) YZ PLANE

FIG. 22. Space diagrams, pattern equations, and sketches of patterns of a single curtain of antennas, each carrying an rms current $I_m/\sqrt{2}$, all currents being in phase.

The equations of the patterns in the other principal planes are obtained in a similar manner. For the xz plane the contribution from a row is

$$\frac{60 I_m}{r} F_0(\theta) F^{4\frac{1}{2}}(\theta)$$

and the contribution from all three rows is

$$\mathcal{E}_{\theta m} = \frac{60 I_m}{r} F_0(\theta) F^{4\frac{1}{2}}(\theta) F^{3\frac{1}{2}}(90°)$$

This result appears to originate at the center of the middle row. The xz-plane plot is shown in Fig. 22B. Likewise, for the yz plane the con-

tribution from a row is

$$\frac{60I_m}{r} F_0(90°)F^{4\frac{1}{2}}(90°)$$

and the contribution for three rows is

$$\mathcal{E}_{\theta m} = \frac{60I_m}{r} F_0(90°)F^{4\frac{1}{2}}(90°)F^{3\frac{1}{2}}(\alpha)$$

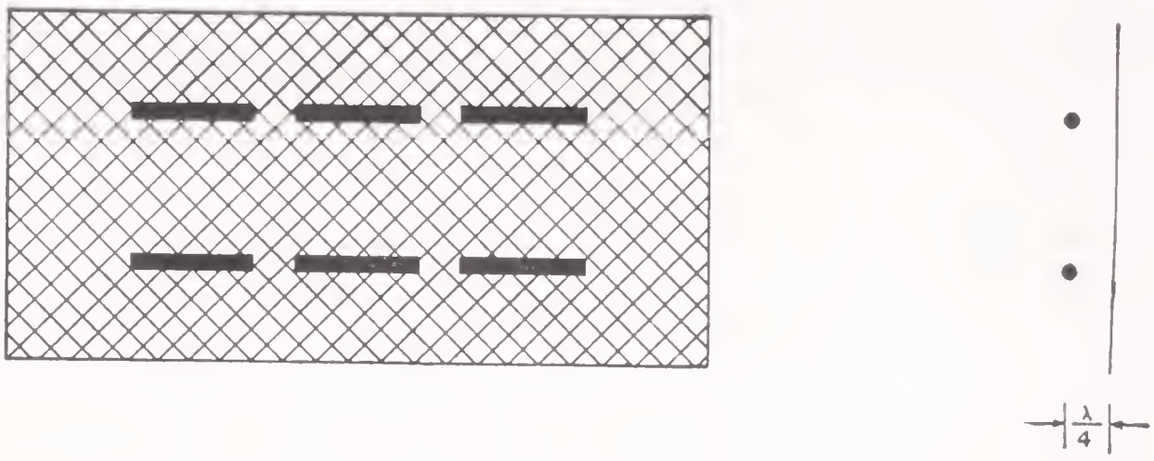
The apparent source of this radiation is at the center of the middle row. The plot is shown in Fig. 22C.

An examination of the patterns for the three planes shows (1) that the width of the array controls the shape of the pattern in the xz plane, (2) that the height of the array controls the shape of the pattern in the yz plane, (3) that the line of principal radiation is along the z axis or broadside to the array, and (4) that relatively small minor lobes exist in the xy plane.

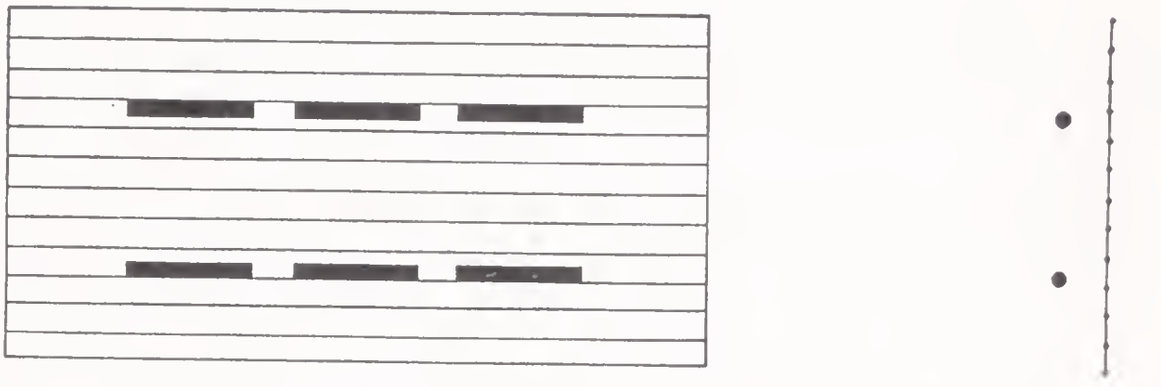
9. Elimination of Back Radiation from Broadside Arrays. It may be noted from the examples in the preceding article that a single curtain of half-wave antennas gives broadside radiation in equal amounts on the two sides of the curtain. This feature of having two opposite directions of principal radiation is undesirable in radar. One of the two main lobes may be eliminated either by adding a second curtain of antennas behind the first or by placing a conducting screen approximately a quarter wave behind the original curtain. This conducting screen could be a solid sheet of metal, but more often it consists of a screen or grid structure, as indicated in Fig. 23. The screen in Fig. 23*A* allows the antennas to be placed either vertically or horizontally, whereas the screens in Figs. 23*B* and *C* are effective only if the antennas are parallel to the grid wires. In Fig. 23*C* the grid wires are spaced so that the angles subtended by adjacent grid wires are equal. The most economical use of material is thus achieved.

Reflecting screens are usually preferable to the two-curtain method of eliminating back radiation. Screens are not critical either as to size or as to spacing from the driven curtain, and they give excellent results if the screen is made only a little larger than the driven curtain.

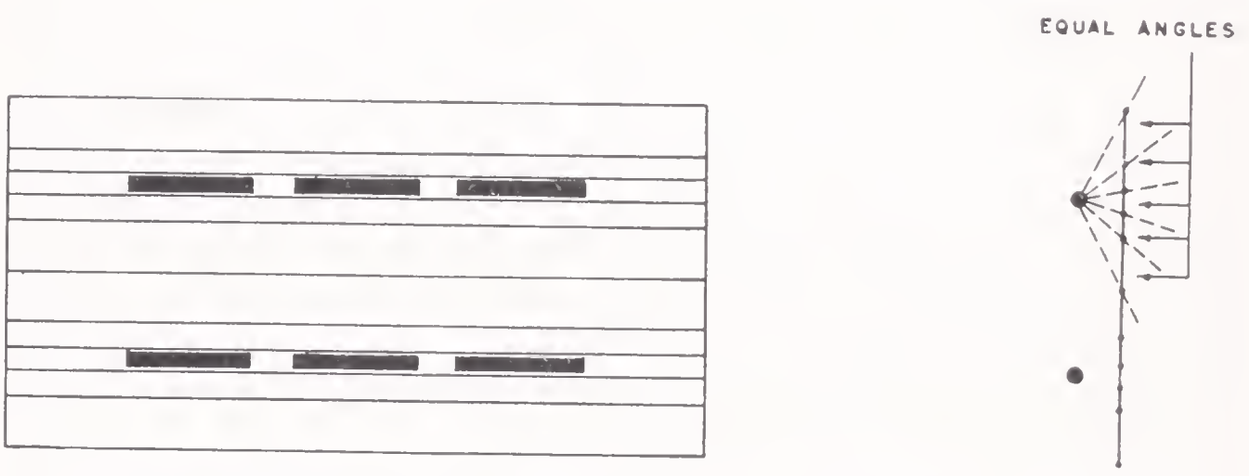
The presence of a screen behind a broadside curtain has two effects: It essentially eliminates back radiation by reflection, and it sharpens the forward pattern. The effect on the forward pattern can be rather accurately determined by use of the principal of images. Figure 24 shows that a real antenna in the presence of a reflecting screen sends radiation to an arbitrary point P along two paths—one direct and one reflected from the screen. The problem of combining the fields at P



(A) MESH SCREEN



(B) EQUALLY SPACED PARALLEL-WIRE SCREEN



(C) UNEQUALLY SPACED PARALLEL-WIRE SCREEN

FIG. 23. Examples of screen reflectors.

due to waves on the two paths reduces to a consideration of the phase of arrival because of the difference in the lengths of the paths. A consideration of the geometry shows that the path length of the reflected ray is always the same as though this ray had originated at an image antenna located as far behind the screen as the real antenna is in front of the screen. Furthermore, there is a reversal in phase, upon reflec-

tion from the screen, which introduces an additional 180-deg delay in the reflected ray. This latter effect can be simulated by considering image and real antennas to be fed 180 deg out of phase. Then, as far as the forward radiation is concerned, the effect of the real antenna with its screen reflector is exactly represented by the real antenna and an image antenna which is excited to the same extent in exactly opposite phase and which is located as far behind the screen as the real antenna is in front of it. It should be noted that the image configuration is not

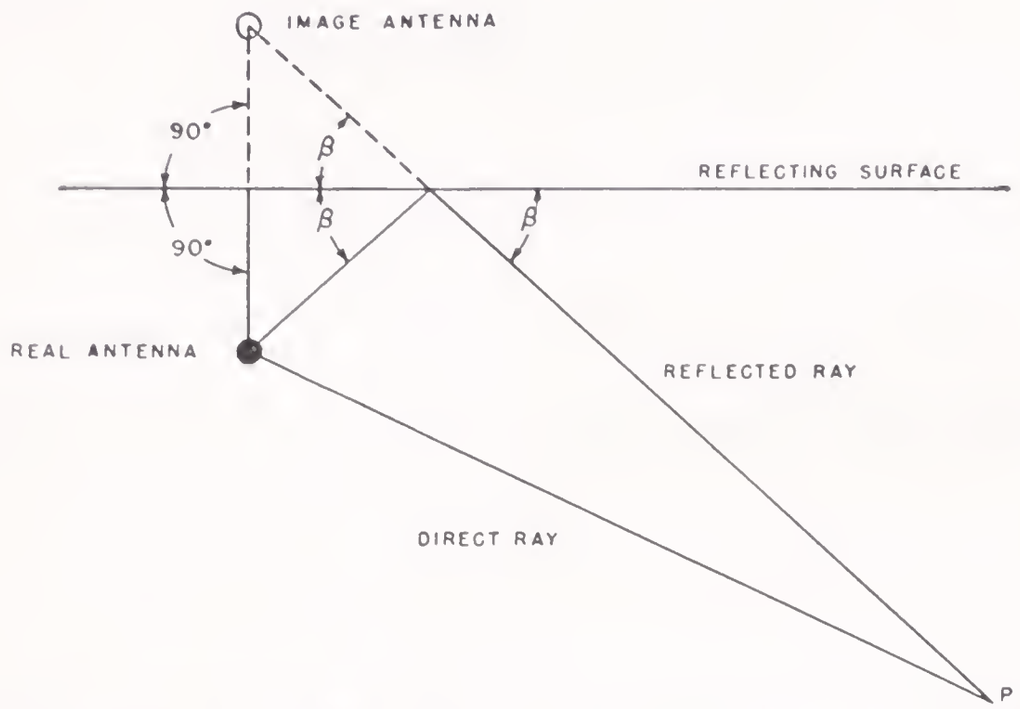


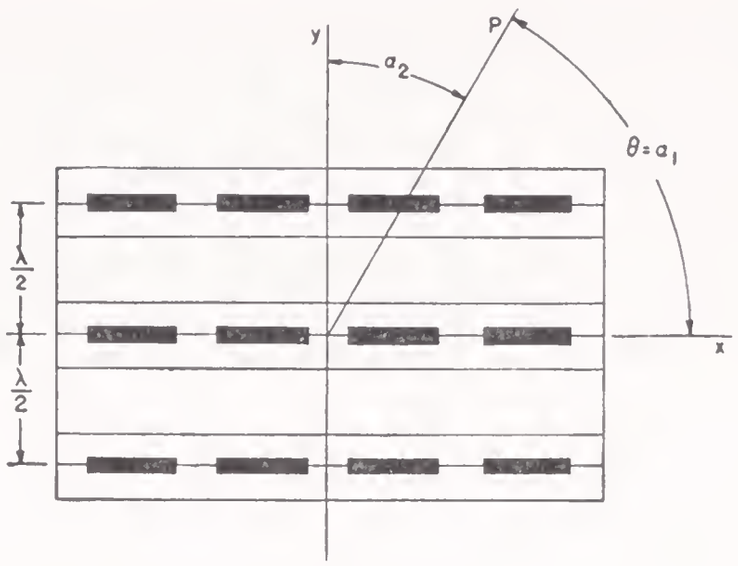
FIG. 24. Real antenna and image antenna which give the same pattern in the forward direction as the real antenna and the reflecting surface.

to be used in obtaining the backward radiation, which is assumed to be zero.

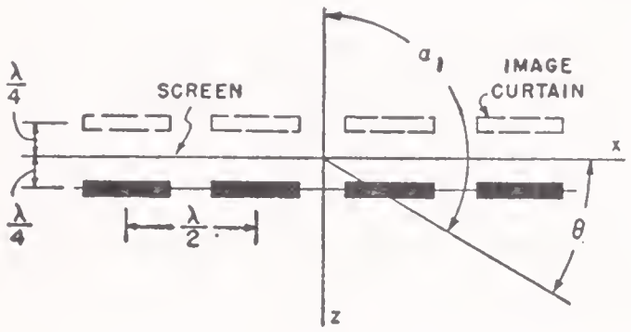
On the assumption that a screen is located $\lambda/4$ from a broadside curtain, the image curtain is $\lambda/2$ from the real curtain and 180 deg out of phase with it. The two curtains can then be added by introducing the factor $G_{\frac{1}{2}}^{2\frac{1}{2}}$ into the pattern equation, with the reservation that this result is correct only for the forward pattern. If the screen is $3\lambda/16$ behind the broadside curtain, the factor is $G_{\frac{1}{2}}^{2\frac{3}{8}}$, and so on.

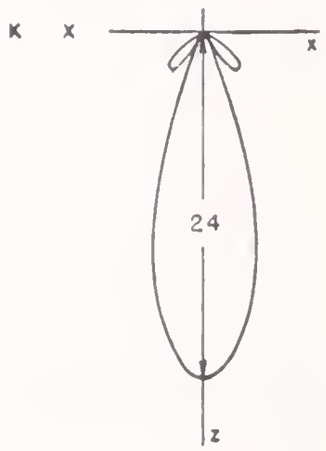
In order to illustrate the effect of a reflecting screen on the radiation pattern of a broadside array, again consider the array of Fig. 22, and assume that a screen is located a quarter wavelength behind the elements. The array and image elements are illustrated in Fig. 25. The pattern equations are the same as those of Fig. 22, except that each equation is multiplied by a $G_{\frac{1}{2}}^{2\frac{1}{2}}$ factor to take into account the effect of the image curtain.

In the xy plane, the angle between a radial line from the center of the array to point P and a line through the centers of front and rear curtain

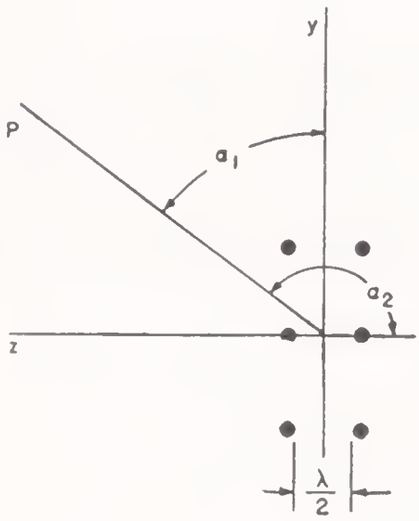


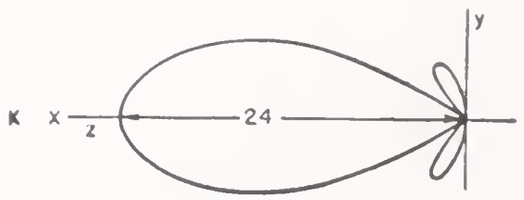
(A) XY PLANE





(B) XZ PLANE





(C) YZ PLANE

FIG. 25. Space diagrams and radiation patterns of a single curtain of antennas backed by a reflecting screen. All elements carry equal currents, and all elements are excited in phase.

is 90 deg. Thus the G factor for the xy-plane pattern is $G_{1\frac{1}{2}}{}^{2\frac{1}{2}}(90°)$, and the xy-plane pattern is given by

$$\mathcal{E}_{\theta m} = \frac{60I_m}{r} F_0(\theta)F^{4\frac{1}{2}}(\alpha_1)F^{3\frac{1}{2}}(\alpha_2)G_{1\frac{1}{2}}{}^{2\frac{1}{2}}(90°) \tag{19}$$

Since the value of $G_{1\frac{1}{2}}{}^{2\frac{1}{2}}(90°)$ is zero (see Fig. 19), no radiation takes place in the xy plane. This is to be expected, since the field at any point in this plane is the result of two component fields of equal amplitude and 180 deg out of phase, one component field being laid down by the real curtain, the other by the image curtain. In the xz plane, the G

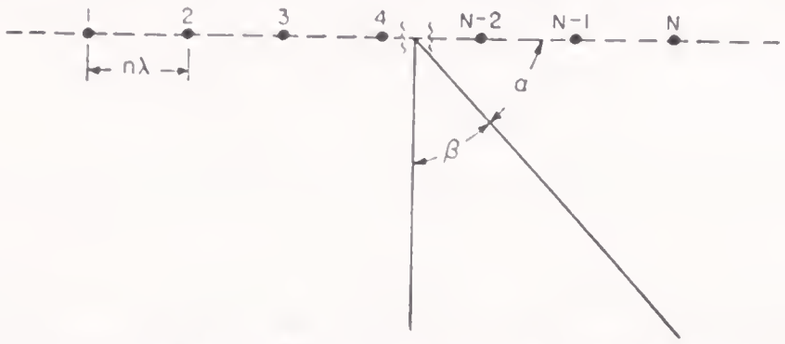


FIG. 26. A broadside array of N identical sources spaced n wavelengths apart.

factor is $G_{1\frac{1}{2}}{}^{2\frac{1}{2}}(\alpha_1)$: thus the xz-plane pattern is expressed as

$$\mathcal{E}_{\theta m} = \frac{60I_m}{r} F_0(\theta)F^{4\frac{1}{2}}(\theta)F^{3\frac{1}{2}}(90°)G_{1\frac{1}{2}}{}^{2\frac{1}{2}}(\alpha_1) \tag{20}$$

For the yz plane

$$\mathcal{E}_{\theta m} = \frac{60I_m}{r} F_0(90°)F^{3\frac{1}{2}}(\alpha_1)F^{4\frac{1}{2}}(90°)G_{1\frac{1}{2}}{}^{2\frac{1}{2}}(\alpha_2) \tag{21}$$

The patterns represented by Eqs. 20 and 21 are plotted in Figs. 25B and C. Observe that no radiation is indicated behind the screen. Calculation of the beam angles of the patterns in Figs. 25 and 22 shows that the presence of the reflecting screen sharpens the main lobe in the forward direction of radiation. The amplitudes of the minor lobes are also reduced, and radiation in the xy plane is completely eliminated.

10. Approximate Method of Determining the Beam Angle of a Broadside Array. The beam angle of an antenna array may be determined from the radiation pattern obtained through use of the proper antenna factors. Because actual computation of the pattern is sometimes laborious, a general expression for the beam angle is often useful. To develop such an expression, first consider N identical sources, each spaced n wavelengths apart and excited in phase (see Fig. 26).

From Eq. 14, p. 918, the electric-field intensity of the array in the plane of the page is

$$\mathcal{E}_{\theta m} = \mathcal{E}_{\theta m1} \frac{\sin\,(N\ 180°n\ \cos\,\alpha)}{\sin\,(180°n\ \cos\,\alpha)} \tag{22}$$

where $\mathcal{E}_{\theta m1}$ is the contribution of one antenna. In terms of the complementary angle β, Eq. 22 becomes, when 180 deg is replaced by π in order to change degrees into radians,

$$\mathcal{E}_{\theta m} = \mathcal{E}_{\theta m1} \frac{\sin (N\pi n \sin \beta)}{\sin (\pi n \sin \beta)} \tag{23}$$

The maximum value of $\mathcal{E}_{\theta m}$ occurs when $\beta = 0$ and is

$$\mathcal{E}_{\theta m_{\max}} = \mathcal{E}_{\theta m1} \times N \tag{24}$$

Dividing Eq. 23 by Eq. 24 yields the per-unit field intensity

$$\frac{\mathcal{E}_{\theta m}}{\mathcal{E}_{\theta m_{\max}}} = \frac{\sin (N\pi n \sin \beta)}{N \sin (\pi n \sin \beta)} \tag{25}$$

The value of β for which $\mathcal{E}_{\theta m}/\mathcal{E}_{\theta m\max} = 0.707$ corresponds to one-half the beam angle. If this value of β is so small that $\pi n \sin \beta$ is itself a small angle, then $\sin (\pi n \sin \beta)$ may be replaced by $(\pi n \sin \beta)$, as the sine of a small angle very nearly equals the angle in radians. Equation 25 then becomes

$$\frac{\mathcal{E}_{\theta m}}{\mathcal{E}_{\theta m_{\max}}} = \frac{\sin (N\pi n \sin \beta)}{N\pi n \sin \beta} = \frac{\sin x}{x} \tag{26}$$

where

$$x = N\pi n \sin \beta$$

The quantity $(\sin x)/x$ has a value of 0.707 when $x = 1.39$ radians, which corresponds to $\sin \beta = 0.443/Nn$. If the angle β in radians is taken as approximately equal to its sine,

$$\beta \approx \frac{0.443}{Nn} \qquad \text{radians}$$

The beam angle, which is 2β, is therefore given in degrees by

$$\text{Beam angle} \approx 2 \times \frac{0.443}{Nn} \times 57.3 \approx \frac{51}{Nn} \qquad \text{deg} \tag{27}$$

The quantity Nn is the number of sources times the spacing of sources in wavelengths, which is the width of the array in wavelengths. Therefore, Eq. 27 states that the beam angle in degrees is approximately equal to 51 divided by the width of the array (in wavelengths) in the plane being considered. This approximation is useful provided Nn is greater than 2, that is, provided the beam angle is less than 25 deg.

11. Antenna Connections for Broadside Arrays. In the preceding articles the currents in the various half-wave antennas forming an array are assumed to be of equal magnitude and to bear a definite phase relation

to one another. The antennas may be fed in phase by making connections of the type indicated in Fig. 27.

The operation may be most readily explained in terms of the voltage at the input ends of the elements. The equal-phase, equal-magnitude-current requirement, when converted to voltage, means that the potentials at the left ends of all antennas should rise and fall together in equal amounts while the potentials at the right ends of all antennas fall and rise in the same amounts. Then a single pair of antennas connected as in Fig. 27*A* will be fed in phase, because the transmission line makes the right end of the left antenna positive at the same time and in the same amount that it makes the left end of the right antenna negative, and vice

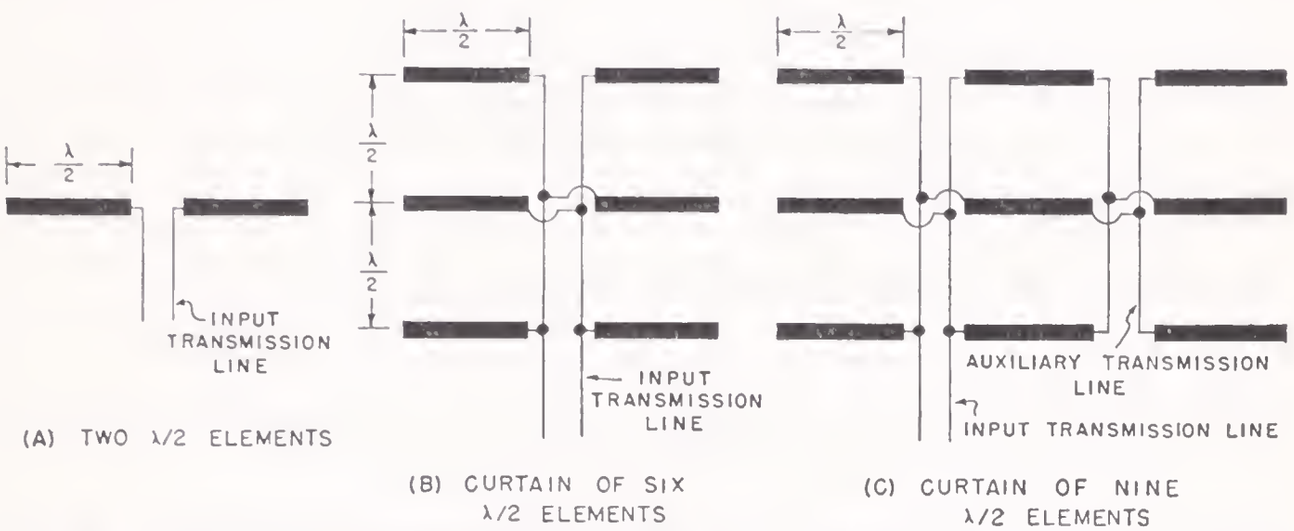


FIG. 27. Connections for end-feeding antennas with in-phase currents of equal magnitude.

versa. This voltage condition is appropriate for equal-phase, equal-magnitude currents.

More than one pair of antennas can be excited in the same manner by connecting the pairs at intervals of $\lambda/2$ along a transmission line, with reversed connections, as indicated in Fig. 27*B*. At intervals of $\lambda/2$ along the transmission line the voltage has opposite phases, whether the line has zero, partial, or complete standing waves along it. Also, the voltage magnitude is exactly the same at these intervals, provided the line is lossless and the loads are either concentrated at the end or distributed at positions which are an integral number of half-wavelengths from the end. However, with loads distributed at half-wavelength intervals, the standing-wave ratio in each half-wave section of line is different; so it is only the voltages at the load points that are identical from section to section.

Figure 27*C* indicates that the antennas do not necessarily have to be grouped in pairs, although such grouping has certain advantages. A second transmission line can be driven by properly connecting it to the antennas which are excited from the main transmission line. This second line can in turn be used to excite additional antennas as shown.

An examination of connections will indicate that the voltage requirements for equal-phase, equal-magnitude currents in all antennas are met.

One feature of Figs. 27*A* and *B* that deserves special comment is that standing waves of appreciable amplitude may exist on the transmission line within the antenna array. The various half-wave sections of line can, in fact, be regarded as pairs of half-wave antennas which are excited to various degrees. These radiate just the same as do the other antennas in the array, but because they occur in pairs which are close together and which are excited to the same extent and in exactly opposite phases, their effect on the pattern is small.

Although the common method of exciting broadside antenna arrays in radar is to use voltage-fed elements, other excitation systems are also in

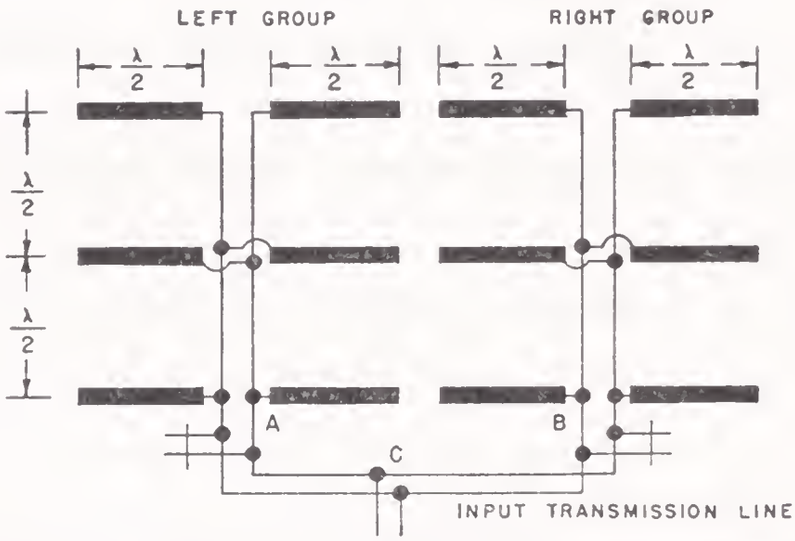


FIG. 28. Two groups of antennas, both groups having equal-phase, equal-magnitude currents, but excited out of phase.

use. Current-fed elements, for example, are used, but their low input impedance prevents them from being well suited to arrays having many elements unless special impedance-adjusting arrangements are introduced. The method of feeding a number of current-fed elements is to arrange them at intervals of $\lambda/2$ along a transmission line, with reversed connections at alternate elements as with the voltage-fed system. But it is not possible with the current-fed system to excite auxiliary transmission lines as in Fig. 27*C*.

An arrangement for exciting two groups of antennas a fixed amount out of phase is indicated in Fig. 28. The six antennas in the left group are all excited in one phase with equal-current magnitudes, and the six antennas in the right group are all excited in another phase with the same current magnitudes. This phasing is accomplished by matching the input impedance of the groups to the line at points *A* and *B* so that no standing waves exist along the line from *A* to *B*. Also, a match is made at point *C* by making the characteristic impedance of the input line half that of the line joining *A* and *B*. Under these circumstances the voltage

everywhere from the input terminals up to points A and B has the same magnitude, but there is a progressive delay in phase. Hence the input voltage to both groups of antennas has the same magnitude (and therefore the currents throughout the groups have the same magnitude), but the phase of the voltage at B lags that at A by the number of degrees that line length CB exceeds line length CA.

12. Parasitic Antennas. Some uhf radar-antenna systems employ parasitic antennas as a means of obtaining directivity and eliminating back radiation. A parasitic antenna is a half-wave antenna which is placed near a similar driven antenna and excited by the field laid down by the driven antenna. The arrangement is shown in Fig. 29.

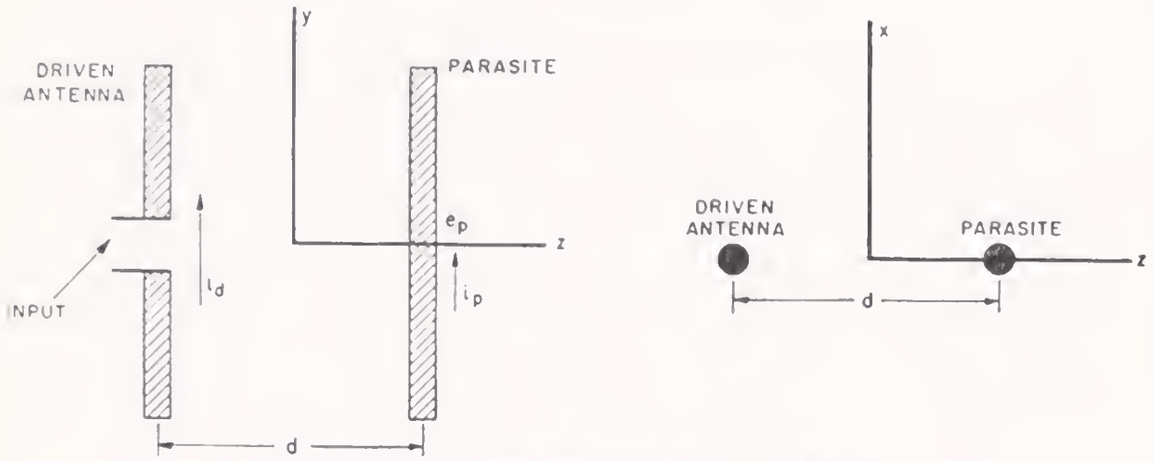


FIG. 29. Driven antenna with parasite.

If a time-varying current flows in the driven antenna, a time-varying voltage is induced in the parasite. The induced voltage is partly due to the induction field set up by the driven-antenna current and partly due to the radiation field. The phase of the induced voltage with respect to the driven-antenna current depends upon the spacing of the elements.

The voltage induced in the parasite may be considered to be equivalent to a voltage source acting in series with the parasite. The magnitude and phase angle of a complex number $E_p = |E_p|\epsilon^{j\phi}$ representing such a series-voltage source are given in Fig. 30 for a unit current at the center of the driven antenna and for various distances d between the antennas. The reference direction for E_p is chosen so that, if the parasite current I_p is assigned the same reference direction as I_d (see Fig. 29), then $I_p = E_p/Z_p$, where Z_p is the self-impedance of the parasite. For example, Fig. 30 shows that if the spacing is approximately 0.15 wavelength, E_p is 180 deg out of phase with I_d. Thus with a 0.15-wavelength spacing and a resonant-length parasite, the two antennas are excited in opposite phase. The data plotted in Fig. 30 are correct for very slender antennas and are approximately correct for thick antennas.

The complex impedance $Z_p = |Z_p|\epsilon^{j\theta_p}$ at the center of the parasite, to which E_p is applied, depends on the length and diameter of the parasite.

For infinitely slender antennas this impedance is a resistance of 73.2 ohms in series with a reactance which depends on the antenna length. When E_p and Z_p are known, the parasitic-antenna current $I_p = E_p/Z_p$ may be calculated, and from knowledge of I_d and I_p the radiation pattern may be determined by the method given in Art. 7.

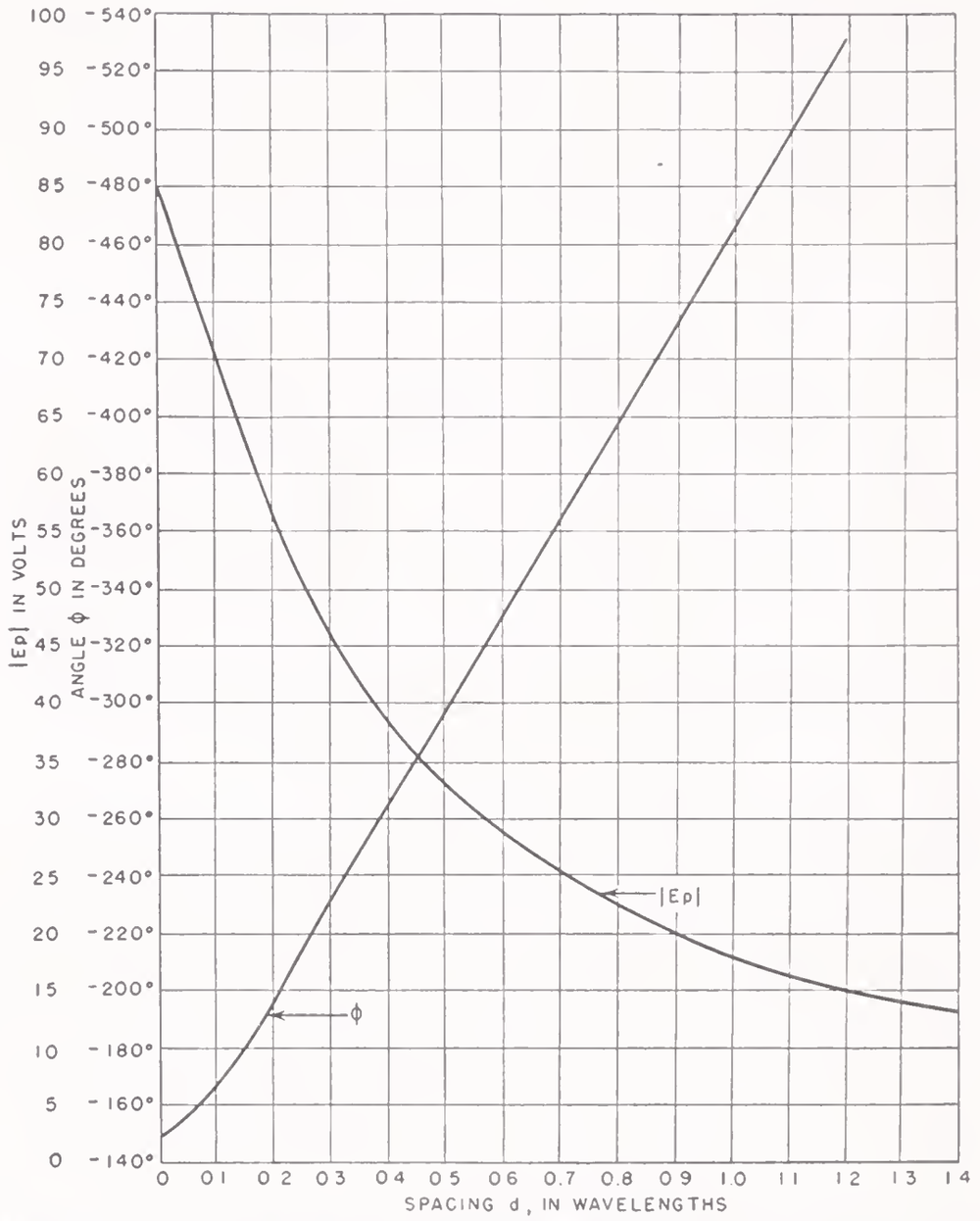


Fig. 30. Magnitude and phase angle of an equivalent voltage source $E_p = |E_p|\epsilon^{j\phi}$ acting at the center of a parasitic antenna, per-unit current in driven antenna.

The patterns obtainable in the magnetic (xz) plane with a single parasitic antenna and driven antenna are shown in Fig. 31 for several antenna spacings and values of θ_p corresponding to several lengths of parasite. These patterns are based upon the assumption that the power input to the driven antenna is constant for each condition. With each pattern is given the resistive component of the input impedance at the center of the driven antenna. When the parasite is placed near the driven antenna, the input impedance of the driven antenna is changed, because

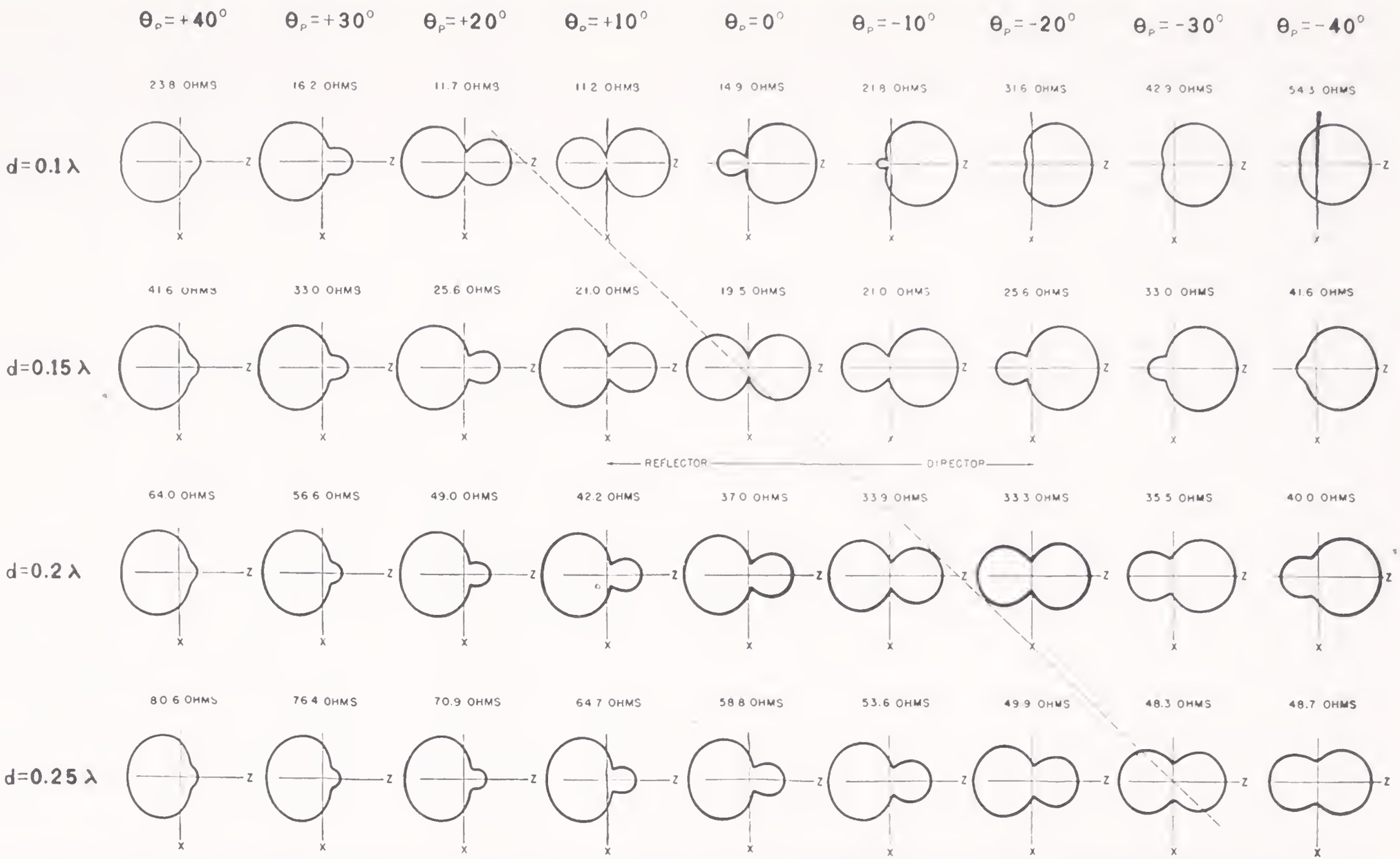


FIG. 31. Patterns obtained in the magnetic (xz) plane of radiation from the antenna system of Fig. 29. The input power is held constant for each condition. Figures above each diagram represent resistive component of input impedance at center of driven antenna.

the current in the parasite causes a voltage to be induced in the driven antenna. The input resistance is usually lowered by the parasite.

Some of the patterns of Fig. 31 have their maximum on the same side of the driven element as the parasite. These are called director patterns because the parasite seems to be pulling the pattern toward itself. Other patterns have a maximum on the opposite side from the parasite. These are called reflector patterns because the parasite seems to be pushing the pattern toward the driven antenna. Good director patterns are to be found where the spacing is 0.1λ and the parasite is shorter than a resonant length, making θ_p about -20 deg. Good reflector patterns occur when the spacing is more nearly 0.2λ and the parasite is longer than a resonant length. Note that it can always be determined whether the parasite acts as a director or as a reflector by remembering that the pattern is pulled toward the element carrying the lagging current.

In order to illustrate the method of determining parasitic-antenna current, let a driven antenna of resonant length be spaced 0.1λ from a parasitic antenna which is shorter than a resonant length by that amount which makes its impedance angle $\theta_p = -20°$, and assume the driven-antenna current to be represented by the complex number $I_d = 2\epsilon^{j0}$. From Fig. 30, the voltage induced in the parasite by a driven-antenna current of unit magnitude has a magnitude of 69.5 volts and lags the driven-antenna current by 166 deg. Hence the complex number E_p standing for the voltage induced in the parasite is

$$E_p = 2\epsilon^{j0} \times 69.5\epsilon^{-j166°} = 139\epsilon^{-j166°}$$

The complex number representing an impedance with a resistive component of 73.2 ohms and an impedance angle of -20 deg is $Z_p = 77.7\epsilon^{-j20°}$. The parasite current is therefore

$$I_p = \frac{E_p}{Z_p} = \frac{139\epsilon^{-j166°}}{77.7\epsilon^{-j20°}} = 1.79\epsilon^{-j146°}$$

That is, a current of 2 amp at the center of the driven antenna causes a current of 1.79 amp to flow in the parasite, and this current lags the driven-antenna current by 146 deg. The corresponding radiation pattern in the xz plane is the seventh diagram from the left in the top row of Fig. 31. The pattern for the yz plane is the xz-plane pattern multiplied by the $F_0(\theta)$ factor.

An increase in directivity and a decrease in back radiation may be obtained by placing parasitic antennas on both sides of the driven antenna. Through proper spacing of the elements, one set of parasites can be made to serve as directors to push on the driven-antenna pattern while the other set serves as reflectors to push against the pattern. An

example of such an array, called a Yagi antenna, is given in Fig. 32. Usually only one reflector is used, because the field behind a reflector is so small that additional reflectors have but little effect. In radar between one and five directors are used. These antennas give patterns having beam angles down to about 40 deg with little back radiation. Smaller beam angles can be obtained through the use of more directors.

The spacings between elements in a Yagi antenna are not uniform. The object is to have the phase angle between the currents in adjacent elements equal to the electrical distance between the elements. Then components of radiation from the various elements arrive in phase along the forward direction but arrive out of phase in various amounts in the

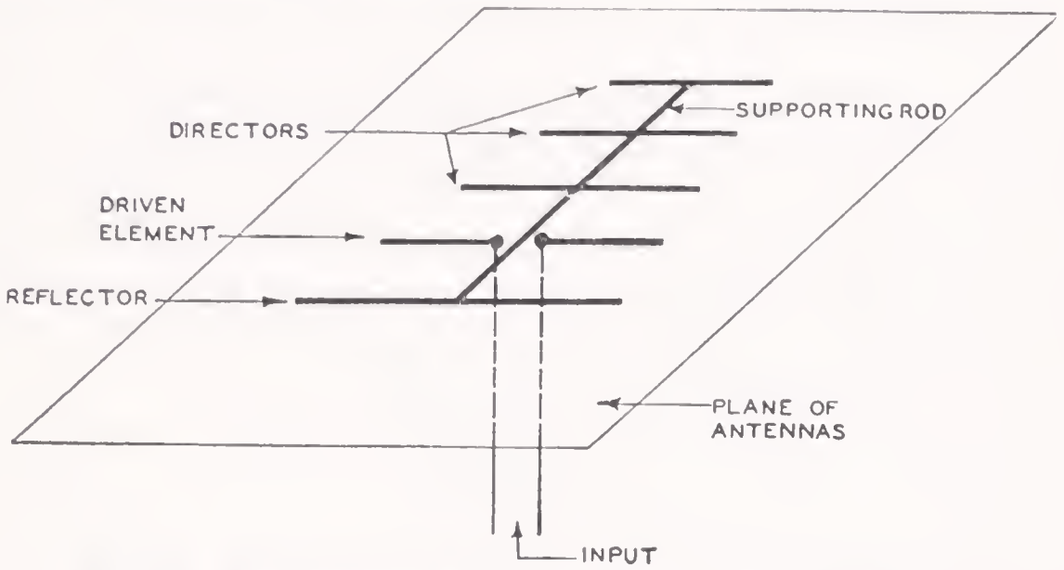


FIG. 32. Yagi antenna with three directors and one reflector.

other directions. Proper spacings and lengths are determined experimentally. Calculations are not profitable with more than two or three elements because the array of dipoles has to be treated as a multielement coupled circuit, each element being coupled to all others.

The parasitic elements of the Yagi antenna are usually welded to a conducting rod or tube at their centers, as indicated in Fig. 32. A conducting support of this kind does not interfere with the operation of the antenna. The driven element, because it has to be fed at the center, cannot be welded to the conducting support.

The input impedances of any Yagi antennas are below the 50 or 60 ohms usually desired, by a factor of perhaps 4 or 5. This impedance can be increased by shortening the driven element and then placing an inductive stub in parallel with the input terminals, as explained in Art. 2. Another method makes use of a *folded dipole* as the driven element. A folded dipole is a center-fed half-wave antenna with another half-wave antenna close to it and joined to the first at the ends, as indicated in Fig. 33. The folded dipole behaves as two parallel dipoles carrying equal currents in the same phase with similar reference directions. Because the

voltage at the ends of the two dipoles is held the same by the end connections, the distribution of voltage along the two antennas is the same. Since the dipoles are alike, equal currents accompany the equal voltages.

If the average power inputs to simple and folded dipoles are compared under the condition of equal magnitudes of input current, the power input to the folded dipole exceeds that to simple dipole by a factor very nearly to 4. Thus a folded dipole has an input resistance of approximately *four* times that of the simple dipole.

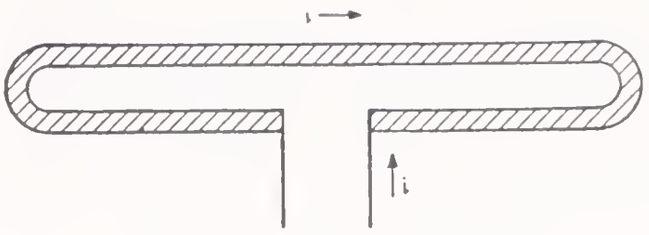


Fig. 33. Folded dipole.

13. Antennas with Parabolic Reflectors. The directional characteristics of antennas are closely related to the shape of the constant-phase surfaces in the electric-field patterns near the antennas. If all the elements of a large broadside array are excited in phase, the phase of the electric-field intensity is nearly the same at all points in a plane a short distance in front of the array. Thus a *plane* constant-phase surface is created having dimensions comparable to the dimensions of the array. The directivity of the array in a given plane depends upon the dimension (relative to a wavelength) of the array in that plane (see Art. 10); consequently the directivity may be considered to depend upon the dimensions in wavelengths of the plane constant-phase surface.

Any antenna that produces a plane constant-phase surface of the same dimensions as the constant-phase surface of the broadside array has essentially the same directivity as the array. In the microwave region, a plane constant-phase surface may be produced simply by placing a parabolic reflector behind a single dipole.

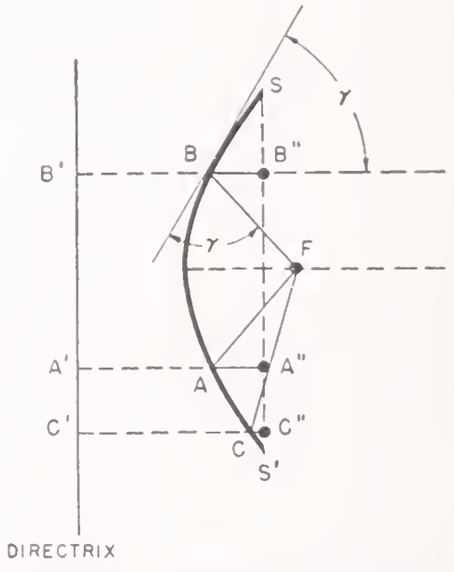


Fig. 34. Diagram for illustrating the geometric properties of a parabola.

Two geometrical properties of parabolas are important for demonstrating the existence of a constant-phase surface. First, a parabola is by definition the locus of points which are as far from a fixed point called the focus as from a fixed line called the directrix. Thus in Fig. 34, lengths AA' and AF are equal, BB' and BF are equal, and so on. Second, a line drawn tangent to a parabola at any point (such as point B in Fig. 34) makes equal angles with a line drawn from this point back to the focus and a line drawn from this point parallel to the axis of the parabola. When a point source of radiation is placed at the focus, energy is emitted

in various directions. This energy strikes the parabolic surface at points such as A, B, and C and is reflected in a direction parallel to the axis because of the second property. The first property predicts that the phase change occurring between the focus and points A'', B'', and C'' on surface SS' is the same for each ray of the emitted wave, the phase change being equal to the distance, in electrical degrees, from the directrix to the surface SS' plus 180 deg, due to the phase reversal upon reflection. Thus the field reflected from the parabola has the same phase at all points in a plane across the mouth of the parabola. The field radiated forward by the point source tends to upset this constant-phase surface, but this effect is usually minimized through the use of sources which radiate appreciably only toward the reflector.

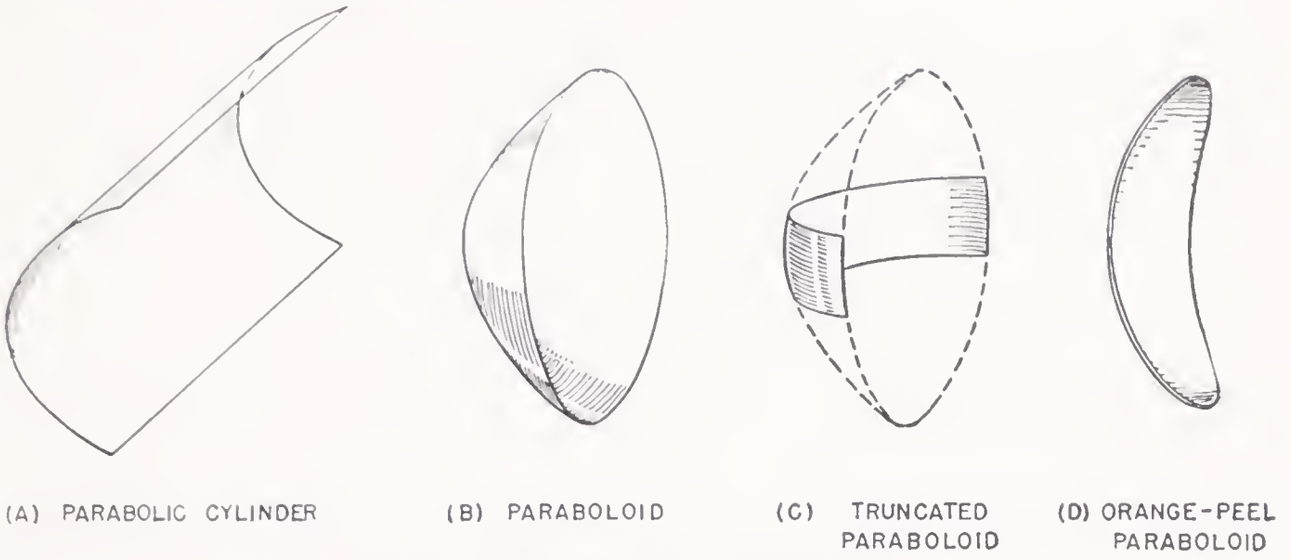


FIG. 35. Typical parabolic reflectors.

Four forms of parabolic reflectors often used in radar are illustrated in Fig. 35. The parabolic cylinder of Fig. 35*A* is illuminated by a row of small antennas along its focal line. Thus directivity in a plane containing the focal line is obtained because of the length of the illuminating source, as in a broadside array. Directivity in a plane perpendicular to the focal line results from the parabolic shape of the reflector. Each paraboloid is illuminated by an approximate point source at its focus, and directivity in any plane results from the shape of the reflector. The reflectors are usually fabricated from solid or perforated sheet metal with appropriate metal support or are made as a screen of properly shaped rods or bars. The screen has the advantages and polarization limitations mentioned in Art. 9.

Consider first the radiation pattern of the parabolic cylinder shown in Fig. 36. The parabolic cylinder is illuminated by a line of half-wave antennas which are fed in phase with equal currents and backed by a line of parasitic reflectors spaced about a quarter wavelength from the driven

antennas. The focal line of the cylinder coincides with the center of illumination—the line midway between the driven antennas and the parasitic antennas.

The uniformity of illumination in the magnetic plane (the yz plane in Fig. 36) depends upon the angle α. The radiation pattern of the illuminating array is determined by the factor $G_{1/4}^{2/4}$ (see Fig. 19). Calculation shows that, for $\alpha = 45°$, the intensity of the field radiated toward the edge of the reflector is 97.5 per cent of that radiated toward the center of the reflector. The illumination of the constant-phase surface at the mouth of the parabola, however, is less uniform than the radiation pattern of the source. Energy radiated from the source in a small-angle wedge and reflected from the center of the parabolic cylinder is spread

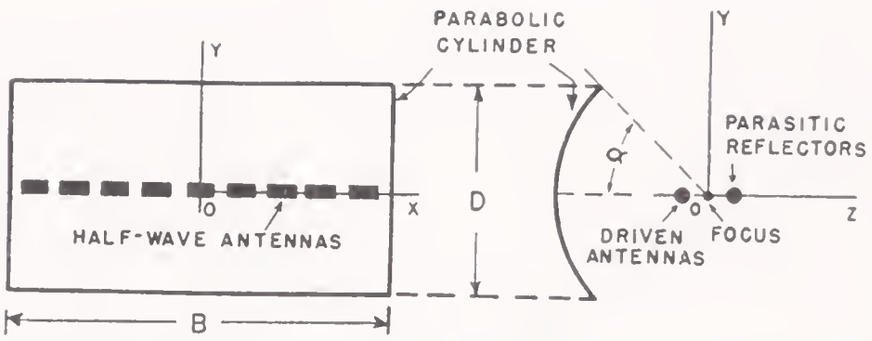


FIG. 36. Parabolic cylinder illuminated by a row of half-wave antennas, each antenna backed by a parasitic reflector.

over a smaller area of the constant-phase surface than is energy in a similar wedge reflected from a part of the cylinder near its edge. From the geometry of the parabola, this effect alone can be shown to reduce the electric-field intensity at the edge of the constant-phase surface to 85 per cent of its value at the center. The supports and feed line for the illuminating array have an additional effect upon the distribution of illumination; generally they tend to reduce the illumination at the center. For simplicity, uniform illumination in the magnetic plane is assumed in the present calculation.

In the xz, or electric, plane the illumination of the parabolic cylinder is nearly uniform. The radiation pattern of the long line of dipoles is highly directive for points at large distances from the array; nevertheless, if the antennas of the feed are closely spaced, the near-by reflector is illuminated almost uniformly, each antenna supplying most of the illumination for parts of the reflector near it.

Thus the electric field has the same phase and nearly uniform amplitude over the entire plane surface across the mouth of the reflector. This surface may be regarded as a large number of small areas, all radiating independently. If the electric field in a small area of the constant-phase surface is $\mathcal{E}_{im}$ and the area is dA, the magnitude of the contribution made

by the small area at a large distance r from the area is approximately

$$d\mathcal{E}_{\theta m} = \frac{1}{2\lambda r}\,\mathcal{E}_{im}dA\,(1 + \cos\beta) \tag{28}$$

Figure 37 shows an element of the constant-phase surface (in the xy plane) and the angle β, which is measured from a perpendicular to the constant-phase surface. The radiation from a small area is similar to that from a half-wave antenna; the former constant 60 is here replaced by $1/(2\lambda)$, the I_m by $\mathcal{E}_{im}dA$, and the $F_0(\theta)$ factor by $(1 + \cos\beta)$. The pattern from a

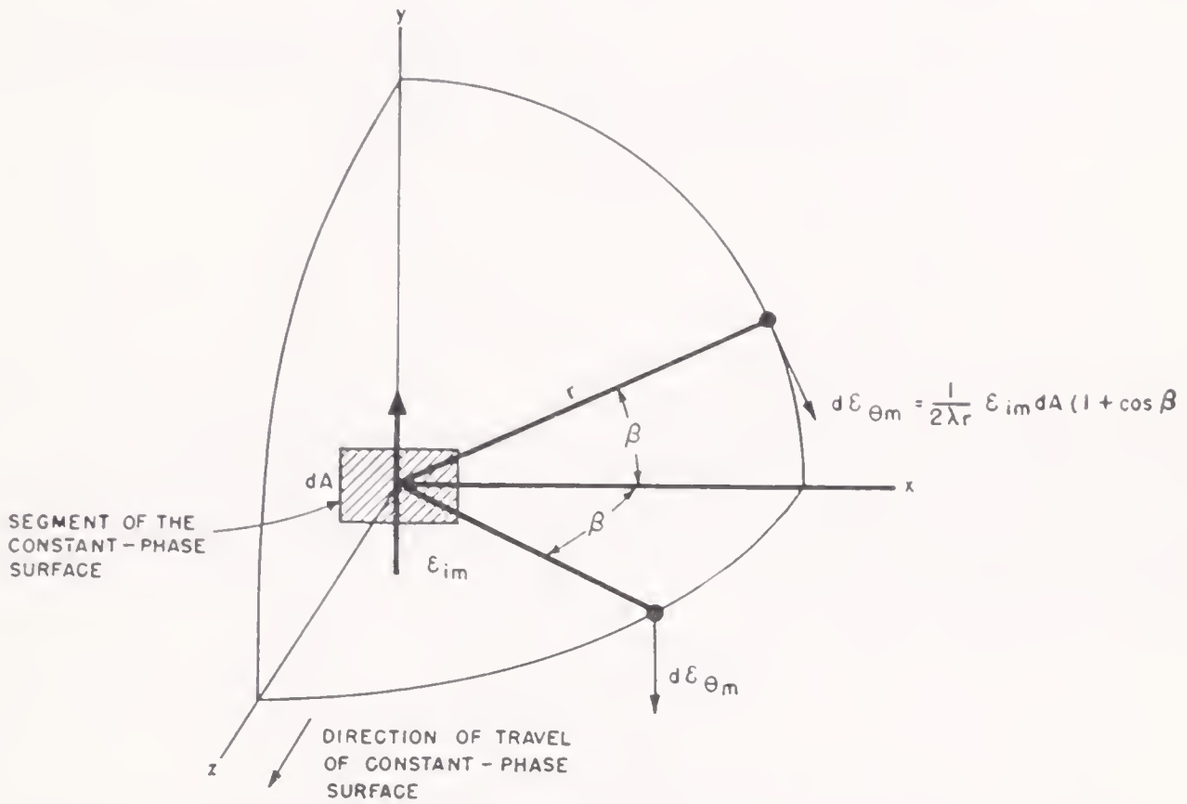


FIG. 37. Element of a constant-phase surface which acts as an independent radiator and lays down a field at a distant point.

single small area is not the symmetrical doughnut shape obtained for the half-wave antenna but has the form of a surface generated by revolving Fig. 38 about the z axis.

The rectangular constant-phase constant-amplitude surface of the parabolic cylinder is drawn in Fig. 39. To determine the pattern in the yz plane, this rectangular area is divided into N small strips parallel to the x axis, each of width D/N. The strips need not be subdivided into squares because all parts of any one strip are at the same distance from a remote point P in the xy plane. Each strip contributes a field at point P of peak value approximately equal to

$$\Delta\mathcal{E}_{\theta m} = \frac{1}{2\lambda r}\,\mathcal{E}_{im}\,\frac{BD}{N}\,(1 + \cos\beta) \tag{29}$$

where BD/N is the area of one strip, corresponding to dA in Eq. 28. An approximation is involved because the strip has a finite width.

Since all the strips radiate in phase and with equal strengths, the contribution from the whole area is an F factor times the contribution from one strip. The F factor is the one which adds N strips spaced $n = D/(N\lambda)$ wavelengths between strips and may be obtained from Eq.

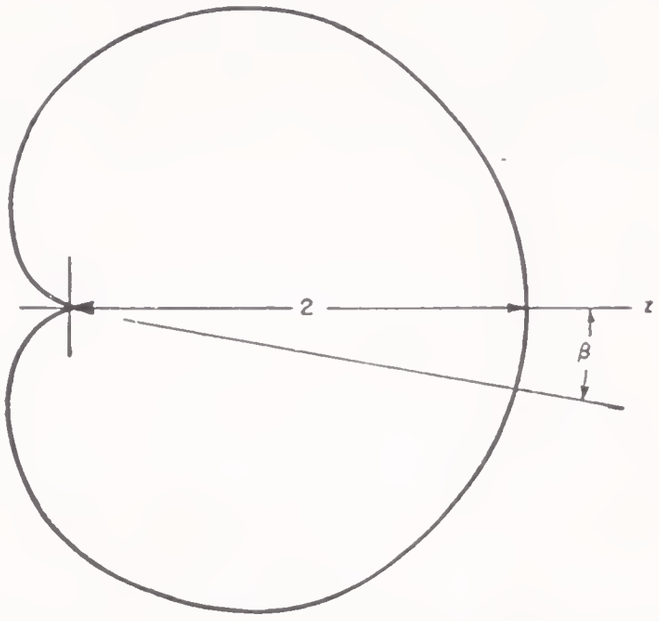


FIG. 38. Plot of $1 + \cos \beta$. The three-dimensional radiation pattern from a small area like that shown in Fig. 37 is obtained by revolving the curve about the z axis.

23, p. 934, by substituting $D/(N\lambda)$ for n. If angles are measured in radians, the factor becomes

$$\frac{\sin\left(\frac{\pi D}{\lambda}\sin\beta\right)}{\sin\left(\frac{\pi D}{N\lambda}\sin\beta\right)}$$

The total field from the constant-phase surface is this factor times the contribution from one strip, or approximately

$$\mathcal{E}_{\theta m} = \frac{1}{2\lambda r}\,\mathcal{E}_{im}\,\frac{BD}{N}\,(1 + \cos\beta)\,\frac{\sin\left(\frac{\pi D}{\lambda}\sin\beta\right)}{\sin\left(\frac{\pi D}{N\lambda}\sin\beta\right)} \tag{30}$$

To remove the approximation from Eqs. 29 and 30, N is made to increase indefinitely, thus making the individual strips infinitely narrow. For such large values of N, the quantity $(\pi D/N\lambda)\sin\beta$ is very small, and the sine of this quantity is equal to the quantity itself. A cancellation of $1/N$ in the numerator and denominator of Eq. 30 is obtained, and the total field is

$$\mathcal{E}_{\theta m} = \frac{1}{2\lambda r}\,\mathcal{E}_{im}BD(1 + \cos\beta)\,\frac{\sin x}{x} \tag{31}$$

where

$$x = \frac{\pi D}{\lambda} \sin \beta \tag{32}$$

The electric field has its maximum value $\mathcal{E}_{\theta m_{\max}}$ when β is zero and equals $(1/2\lambda r)\mathcal{E}_{im}BD \times 2$. The per-unit electric field then becomes

$$\frac{\mathcal{E}_{\theta m}}{\mathcal{E}_{\theta m_{\max}}} = \frac{1 + \cos \beta}{2} \left(\frac{\sin x}{x}\right) \tag{33}$$

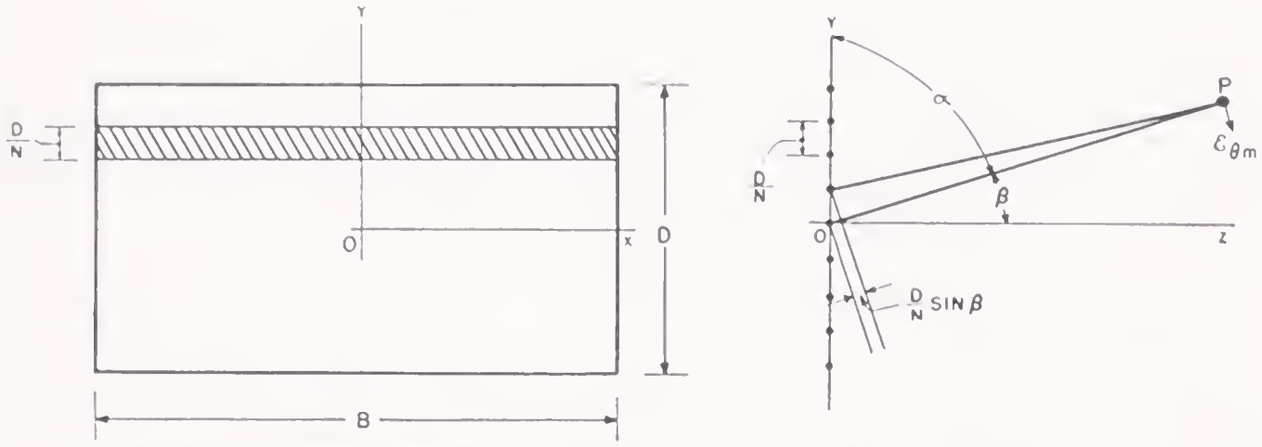


FIG. 39. Constant-phase surface divided into strips which radiate independently. Strips are taken for pattern in yz plane.

The first factor is the directivity of an infinitely narrow strip, and the second factor represents the additional directivity caused by the width of the constant-phase plane. In any radar antenna the first factor is very nearly equal to unity over the range of β where the principal lobes of the pattern occur. Thus the per-unit pattern has approximately the form of $(\sin x)/x$; this function is plotted in Fig. 40.

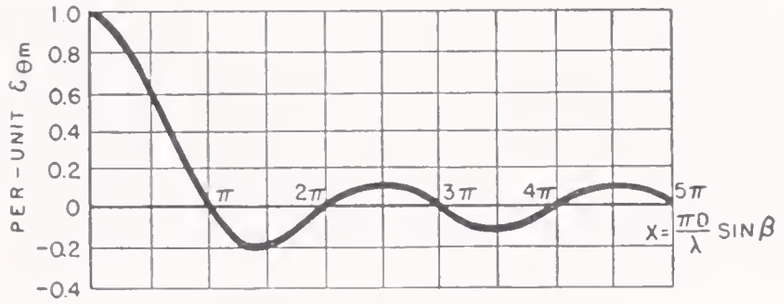


FIG. 40. Curve of $(\sin x)/x$.

The use of Fig. 40 to obtain the pattern of a parabolic cylinder and the general shape of the pattern obtained can both be illustrated by means of a numerical example. Suppose $D = 8\lambda$. The function $(\sin x)/x$ has 0.707 of its maximum value when $x = 0.443\pi$. For this value of x, $\sin \beta = 0.443/8$, according to Eq. 32. Hence $\beta = 3.3°$. If the very small effect of the $(1 + \cos \beta)/2$ factor is neglected, the beam angle of this parabolic cylinder in the yz plane is therefore 6.6 deg. The main lobe goes to zero when $x = \pi$, and an increase in x beyond π produces a change in sign of $\mathcal{E}_{\theta m}$. This change in sign corresponds to a minor lobe. At

$x = 2\pi$ another null occurs, and for $2\pi < x < 3\pi$, another minor lobe appears. A plot of the resulting pattern is given in Fig. 41.

Note, from Fig. 40 and the definition of x, that doubling D decreases the beam angle and increases the number of secondary lobes but does not change the maximum per-unit value of a given lobe. The decrease in beam angle as a result of an increase in the dimension D is analogous to the decrease of beam angle obtained by increasing the width of a broadside array.

The determination of the radiation pattern of the parabolic cylinder in the xz plane is accomplished by the same method as just outlined for the yz plane. Figure 40 applies with x taken to be $(\pi B/\lambda) \sin \beta$.

In the preceding example, the illuminating source was assumed to be an ideal line source, yet a beam angle of over 6 deg was obtained. Note

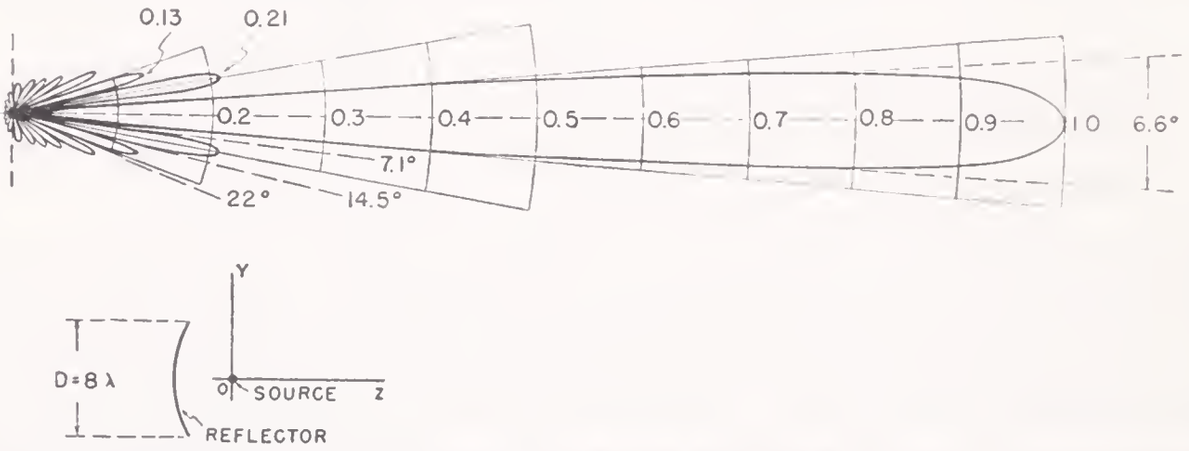


FIG. 41. Radiation pattern in yz plane of parabolic cylinder with uniform illumination of constant-phase surface: $D = 8\lambda$.

that the beam angle of any parabolic antenna is greater than zero *not* because the illuminating source differs from the ideal line source (or point source for a paraboloid) but *because the width (or diameter) of the reflector is not extremely large compared with a wavelength.* To be sure, actual sources are somewhat different from ideal line or point sources, but the effect of the difference is negligible. If radar antennas could be made with diameter-to-wavelength ratios of the same order of magnitude as those in optical systems, then the radar beams would have essentially zero width.

In the constant-phase surface of the parabolic cylinder, all the strips were considered to be alike and to make equal contributions to the field at a distant point. This situation does not always occur with parabolic reflectors for two reasons: (1) The strips sometimes have unequal length, as they do in the circular opening of a paraboloid, and (2) the intensity $\mathcal{E}_{im}$ across the constant-phase surface may not be uniform. Nevertheless, the patterns can always be obtained from curves resembling Fig. 40, plotted in terms of the same variable x.

Curves that specify the radiation patterns of rectangular and circular apertures for two different illuminations of each are given in Fig. 42.

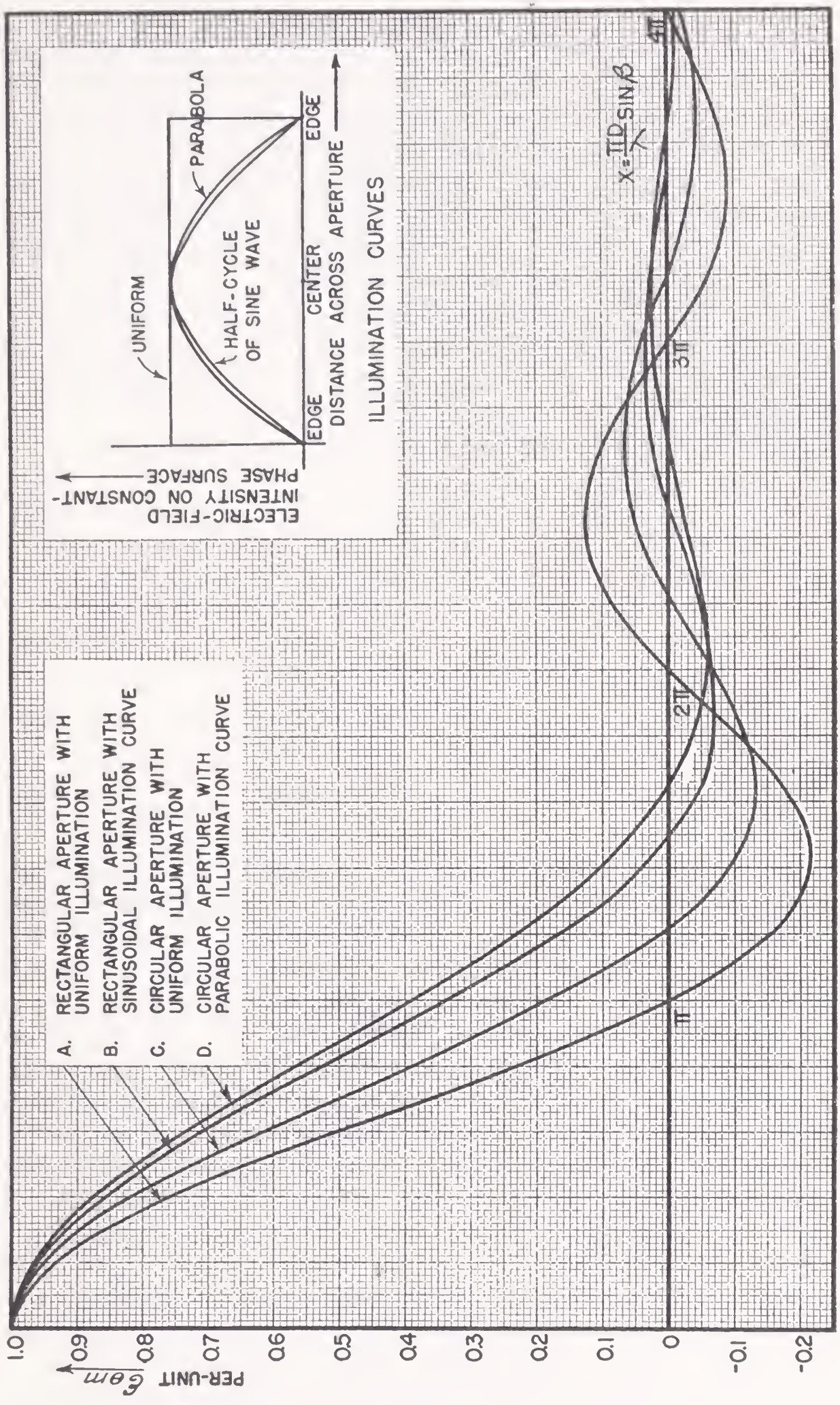


FIG. 42. Curves for determining pattern of parabolic antenna.

Curve A applies to a rectangular aperture with uniform illumination and is the $(\sin x)/x$ curve already discussed. Curve B is for a rectangular aperture with sinusoidal distribution of illumination over the constant-phase surface. Concentrating the illumination in the center causes an increase of beam angle and decrease of minor lobes.

Curves C and D for circular apertures may be used with full paraboloids. Note that the greater width of the center strips of circular openings produces the same general effect on beam angle and minor lobes as does a greater illumination intensity at the center.

Formulas that specify the approximate beam angles of parabolic reflectors may be determined from the curves of Fig. 42 by finding the value of x for which $\mathcal{E}_{\theta m}$ decreases to 0.707 of its maximum value. For example:

For a uniformly illuminated rectangular aperture (curve A):

$$\text{Beam angle} = \frac{51\lambda}{D} \quad \text{deg} \tag{34}$$

For a uniformly illuminated circular aperture (curve C):

$$\text{Beam angle} = \frac{60\lambda}{D} \quad \text{deg} \tag{35}$$

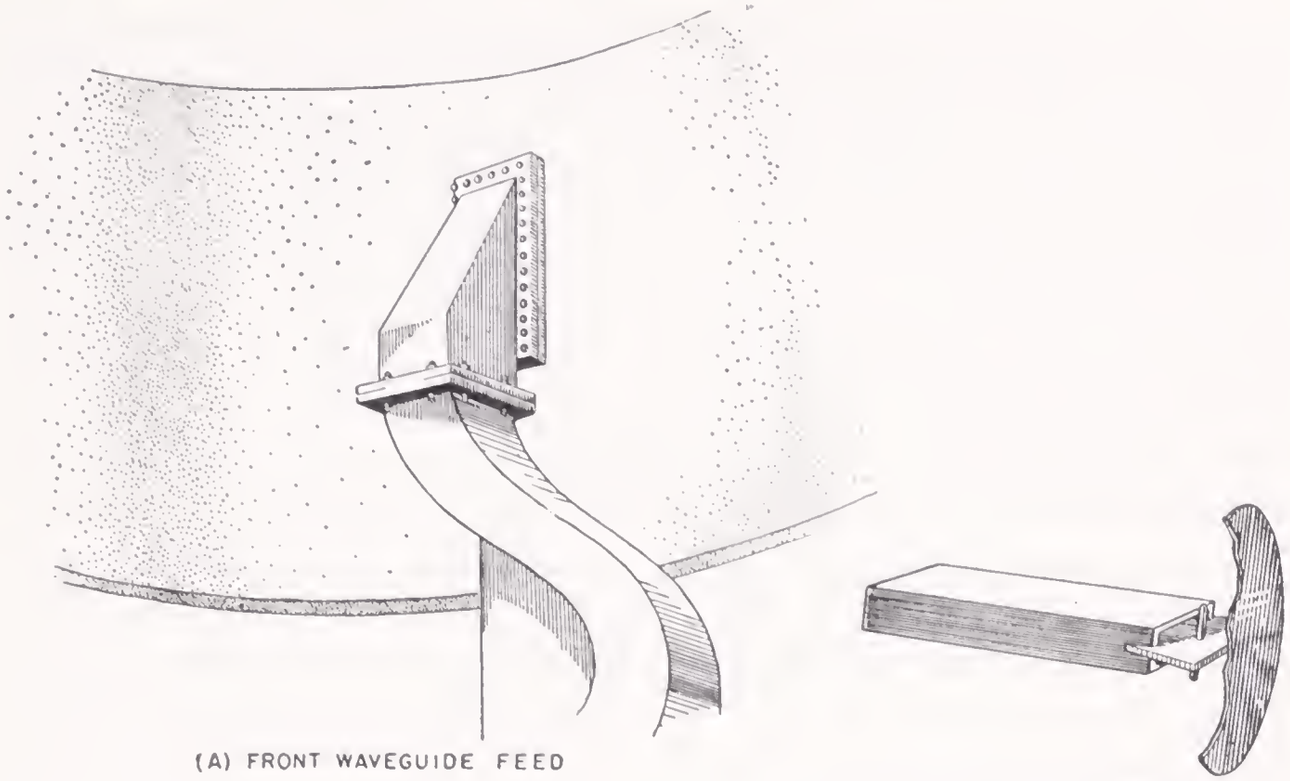
For circular aperture with parabolic distribution of illumination (curve D):

$$\text{Beam angle} = \frac{75\lambda}{D} \quad \text{deg} \tag{36}$$

These equations neglect the factor $(1 + \cos \beta)/2$ in Eq. 33; they are sufficiently accurate for most purposes if the beam angle obtained is less than 20 deg.

14. Feeds for Parabolic Antennas. Antenna feeds, or sources of illumination for parabolic reflectors, may employ either coaxial lines or waveguides and may be classified as front or rear feeds. In a rear feed, the coaxial line or waveguide projects through the reflector, and some device is required at the end of the line or guide to direct energy back toward the reflector. In a front feed, the line or guide approaches the reflector from the front and "sprays" energy directly into the reflector. Figure 43A illustrates a front feed; Figs. 43B and C show rear feeds.

The feed must satisfy two requirements: (1) It must be shaped and located so as to illuminate the reflector in the correct manner. (2) It must terminate the coaxial line or waveguide so that the standing-wave ratio is near unity in the line or guide. To satisfy the first requirement the feed must itself be a moderately directive antenna and must be located so that it does illuminate the entire reflecting surface with a



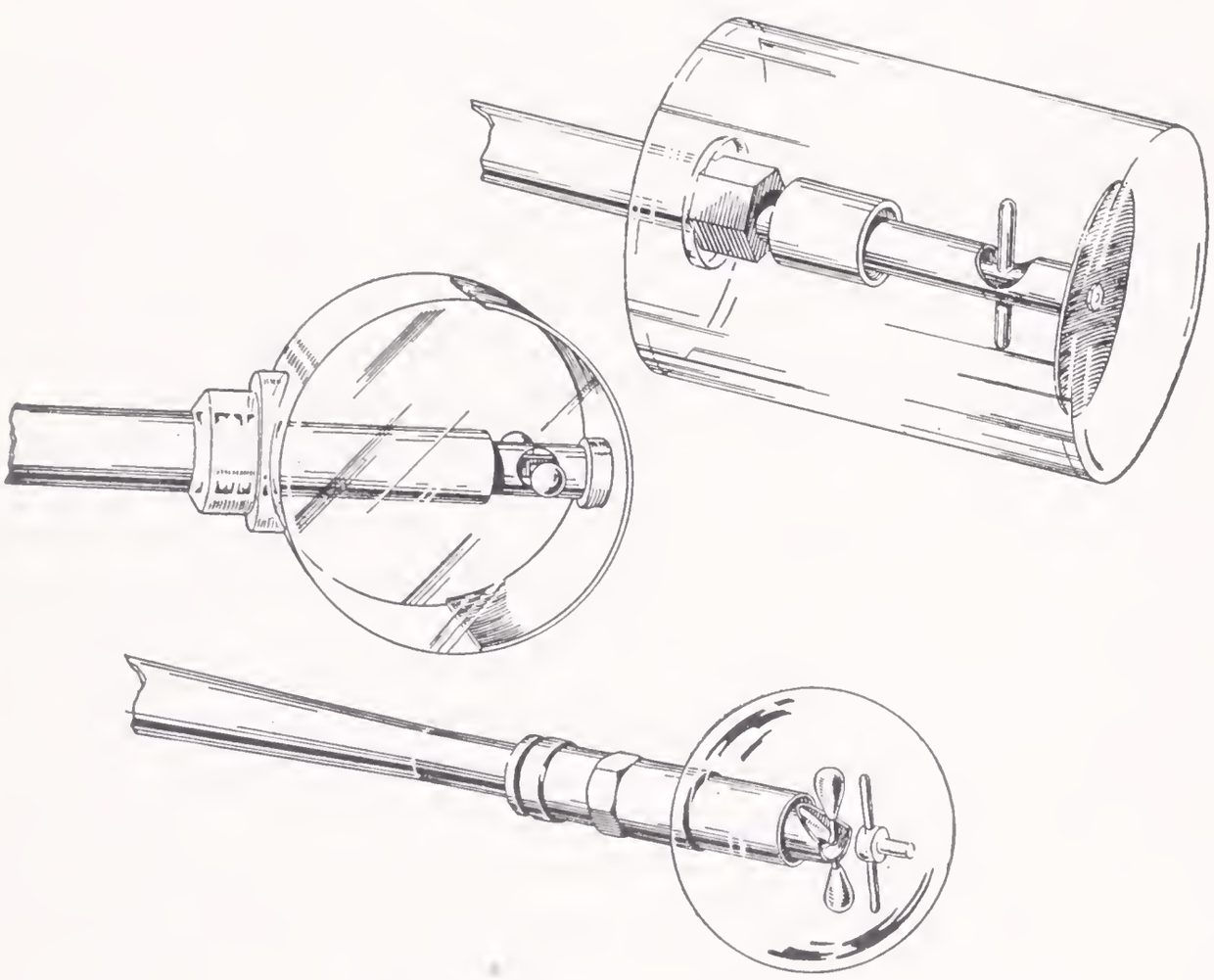


FIG. 43. Examples of antenna feeds.

reasonable degree of uniformity but does not spill too much of the energy over the edges of the reflector. To satisfy the second requirement, impedance-changing devices in or near the feed are required.

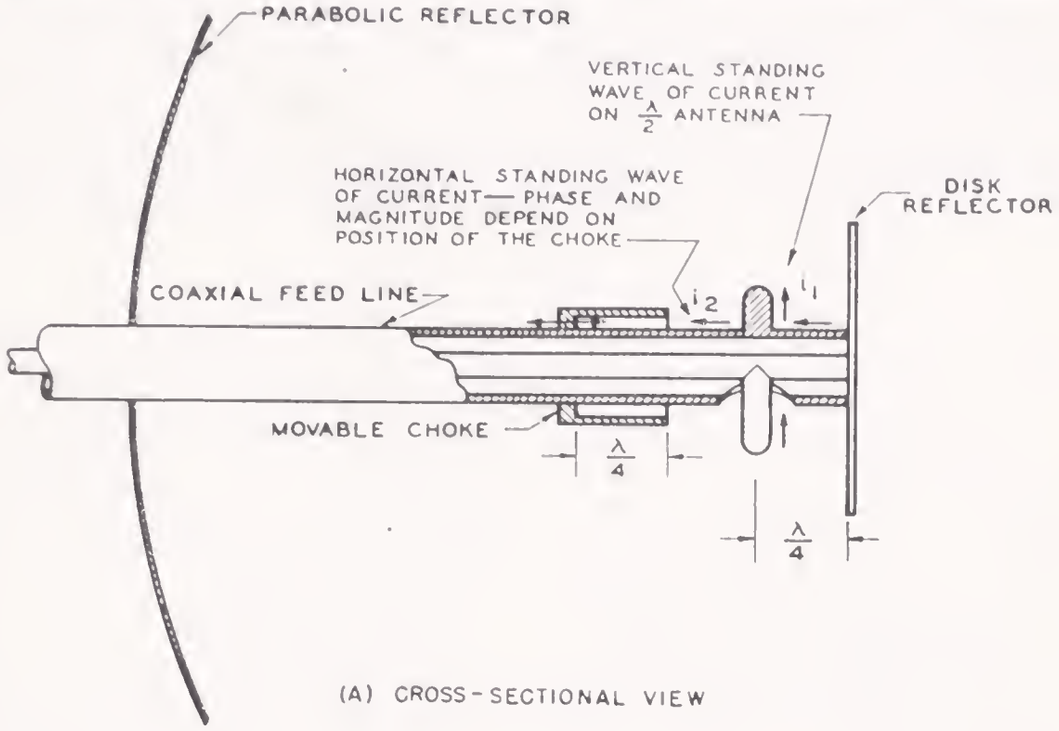
The choice of the focal length of the parabolic reflector depends upon the illumination pattern of the feed. For example, consider a paraboloid of fixed mouth diameter. A short focal length may be used with the feed mounted inside a deep "dish," or a long focal length may be employed with the feed placed far in front of a nearly flat reflector. A short focal length results in nonuniform illumination because of the directivity of the feed; in fact, if some types of feeds are located inside the mouth of the paraboloid, the outer edge of the reflector is illuminated by radiation of reversed phase. Too short a focal length therefore diminishes the useful diameter of the paraboloid. A long focal length provides essentially uniform illumination of the reflector but also allows much of the radiation from the feed to escape reflection. In other words, the over-all beam angle of the main lobe decreases with increase of the focal length, but the gain decreases also, and side and back lobes caused by direct radiation from the feed increase. A practical compromise places the feed a little outside the mouth of the paraboloid.

Adjustment of the focal length of the parabolic reflector does not allow independent control of the illumination in the horizontal and vertical planes. Reflectors like those of Figs. 35*C* and *D* are used when a narrow beam in one plane and a wide beam in a perpendicular plane are desired. *The beam angle of the feed must be large in the plane for which the antenna beam angle is small.* This is necessary so that the reflector may be illuminated fairly uniformly across its wide dimension without excessive energy spilling over its edges in the narrow dimension.

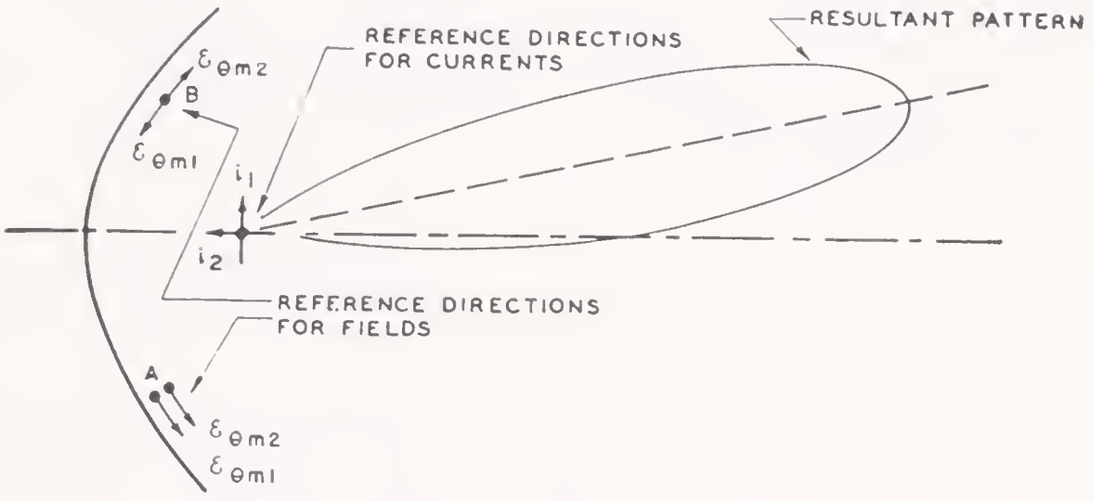
Coaxial-line Feeds. Figure 43*C* shows three typical rear feeds commonly used with full paraboloids in 10-cm radar sets; note that each of the feeds includes a reflector. The parasitic reflector has been largely superseded by the plane disk reflector and the cylindrical reflector. The directivity of the feed is greatest in the electric plane, because in this plane the directivity of the antenna itself (F_0 factor) augments the directivity of the antenna-reflector combination. Consequently, the illumination of the paraboloid is more nearly uniform in the magnetic plane, and if horizontal polarization is used, the beam angle of the antenna system is greater in the horizontal plane than in the vertical plane. In some applications this characteristic is not desirable, but horizontal polarization is nevertheless used for other reasons. The pattern in the magnetic plane is usually about that indicated by curve *C* in Fig. 42; the pattern in the electric plane is close to that specified by curve *D*.

Figure 44*A* shows a coaxial-line feed which provides conical scan when the feed system is rotated. The rotating system is perfectly balanced

mechanically. The half-wave antenna is the main radiating element, the disk minimizes direct radiation forward, and the choke is intended to decouple the radiating element from the outer surface of the feed line. A standing wave is set up on the outside of the coaxial line between the choke and the disk. As a result, two dipoles seem to be radiating—



(A) CROSS-SECTIONAL VIEW



(B) TWO-ANTENNA EQUIVALENT OF PART A

FIG. 44. One form of antenna feed for producing conical scan.

one polarized vertically and one horizontally, as shown in Fig. 44*B*. The magnitude and phase of i_2 with respect to i_1 depend upon the position of the choke. In general the two currents have unequal magnitudes and are out of phase with each other. Also, when the reference directions of current are as shown in Fig. 44*A*, the component electric fields have the same reference direction at points of the parabolic surface below the axis and opposite reference directions at points above the axis. Thus the magnitudes and phases at points such as *A* and *B* in Fig. 44*B* are

different, and consequently the plane of the parabola opening is not a constant-phase surface.

The pattern can be determined by dividing the plane of the parabola opening into strips such that all points on one strip radiate in phase but the fields of the several strips are not in the same phase. The resultant lobe is pulled toward the strips which are lagging in phase, as in arrays. The angle that the axis of the pattern makes with the axis of the parabola can be adjusted by varying the position of the choke.

Front Waveguide Feeds. Figure 43*A* shows a front waveguide feed (called a nozzle feed) illuminating a truncated paraboloid with larger

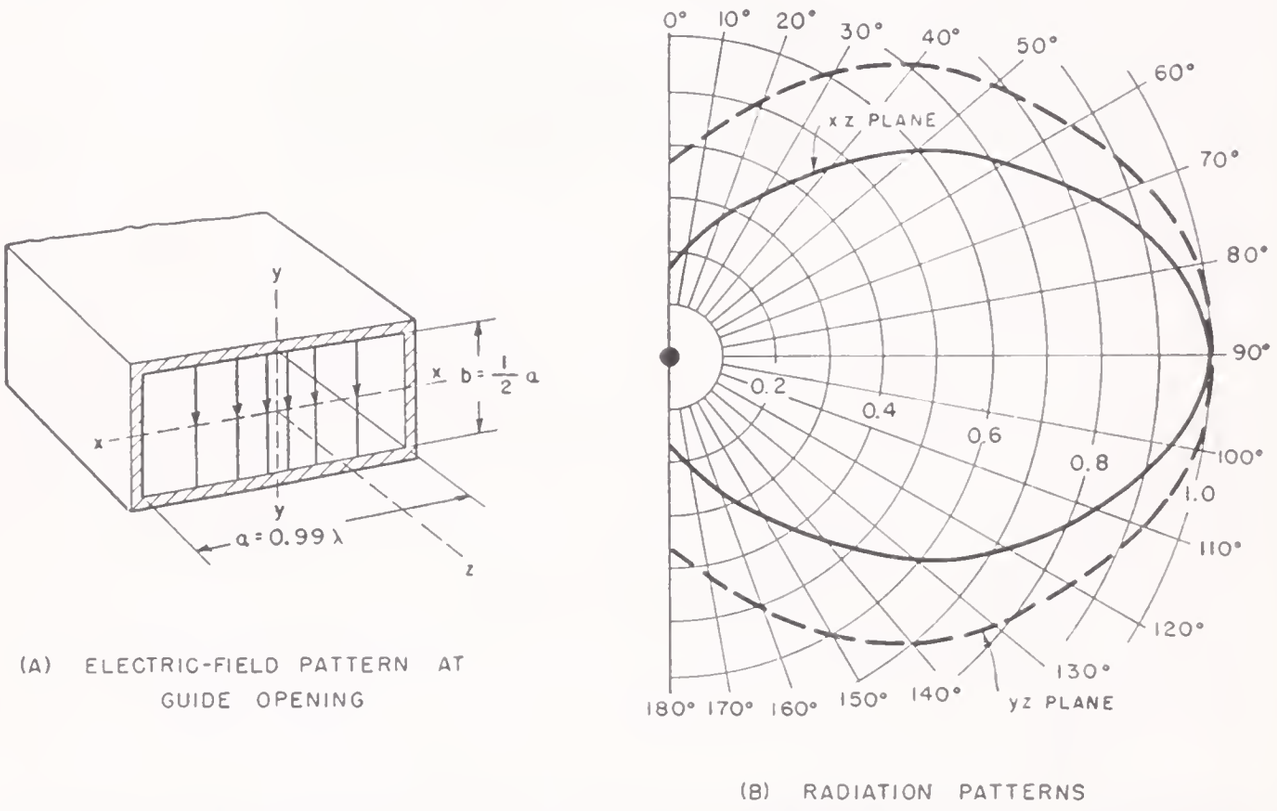


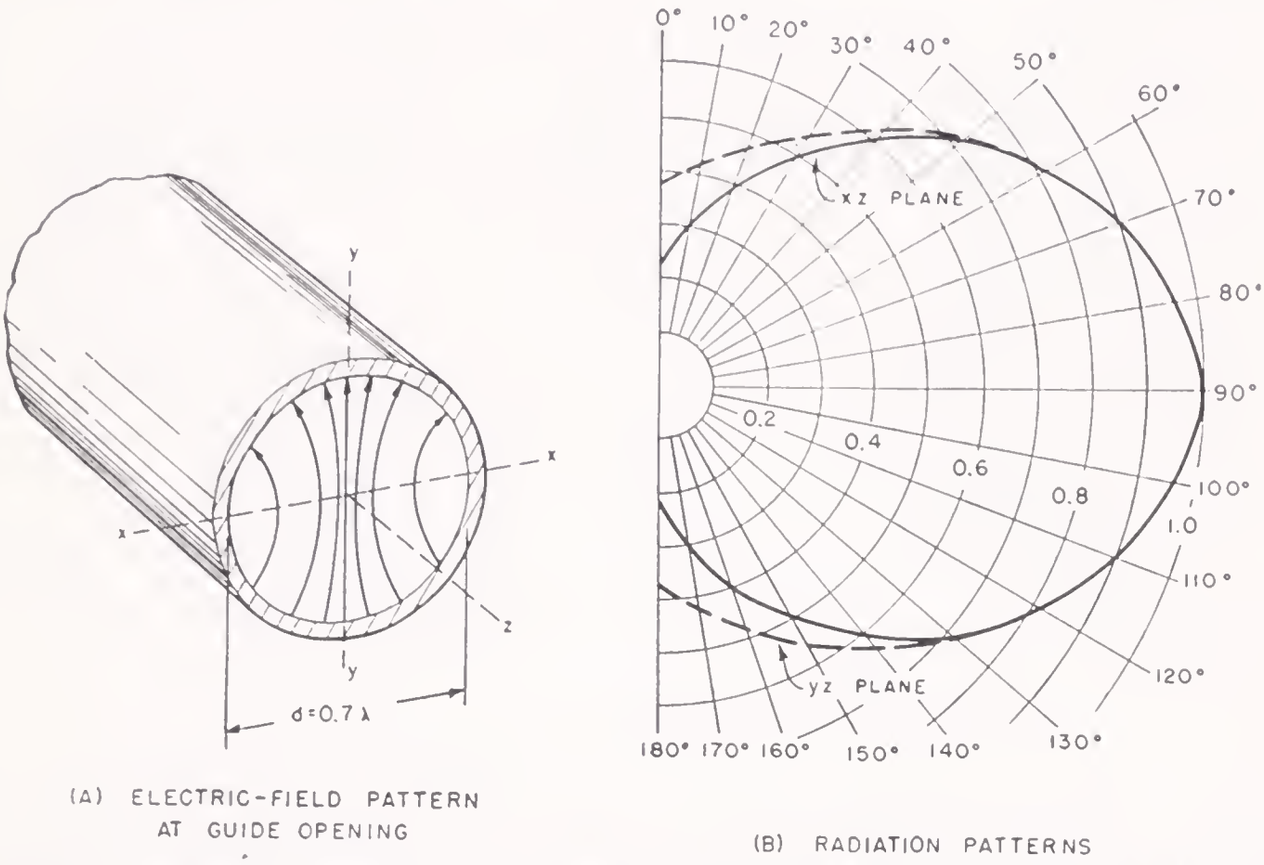
FIG. 45. Radiation patterns from an open-end rectangular waveguide propagating a $TE_{1,0}$ wave.

diameter in the horizontal plane. This combination may be used when a highly directive beam in the horizontal plane is required. The radiation pattern at the mouth of the guide depends on the electric-field configuration at the guide opening. Normally the field configuration is that of a $TE_{1,0}$ wave, as indicated in Fig. 45*A*. Because the electric-field variation is sinusoidal in the x direction, the primary radiation pattern in the xz plane may be determined by reference to curve B in Fig. 42. The $(1 + \cos \beta)/2$ factor cannot be assumed equal to unity here because of the smallness of the aperture. In the yz plane, curve A in Fig. 42 applies. Figure 45*B* shows the patterns of the feed in the xz and yz planes. Orienting the guide with the long dimension parallel to the short dimension of the reflector, as in Fig. 43*A*, provides nearly uniform illumination in the horizontal plane so that maximum directivity

is obtained. In the vertical plane, the shorter reflector dimension produces a larger over-all beam angle, and the directivity of the primary radiation prevents excessive loss of energy.

A circular waveguide propagating a $TE_{1,1}$ wave occasionally serves as a nozzle feed. The polarization of the waveguide field chosen depends upon the polarization of the antenna radiation desired, because there is essentially no difference between the primary directivity in the electric and magnetic planes (see Fig. 46).

The orange-peel paraboloid is shown with its feed in Fig. 47. Assume it to be operated with the longer dimension of the paraboloid vertical.



(A) ELECTRIC-FIELD PATTERN AT GUIDE OPENING

(B) RADIATION PATTERNS

FIG. 46. Radiation pattern from an open-end circular waveguide propagating a $TE_{1,1}$ wave.

The radiation pattern produced is highly directive in the vertical plane, and rather broad in the horizontal plane. The feed is a *sectoral horn*, formed by flaring the wide cross-sectional dimension of a rectangular waveguide (see Fig. 47*B*). The horn is oriented to produce vertical polarization. The flare increases the directivity of the primary field-intensity pattern and reduces the amount of primary radiation that escapes reflection.

As shown in Fig. 47*B*, the constant-phase wavefront at the mouth of the horn is not plane, and therefore the pattern of the primary radiation in the horizontal plane cannot be determined from the curves of Fig. 42. The primary directivity in this plane depends on horn length and flare angle and is generally increased by an increase in horn length. The

over-all beam angle of the antenna in the horizontal plane is large, larger, in fact, than that of a truncated paraboloid of the same width and height and with the same feed. The directivity of the sectoral horn in the vertical plane is small and can be determined by the same method used for the rectangular waveguide nozzle.

The length and flare angle of the sectoral horn not only control the directivity of the radiation which energizes the reflector but also affect the terminating impedance for the waveguide. Flaring the end of a waveguide to form a horn results in a gradual decrease in guide wavelength, so that the wavelength at the mouth of the horn is nearly the

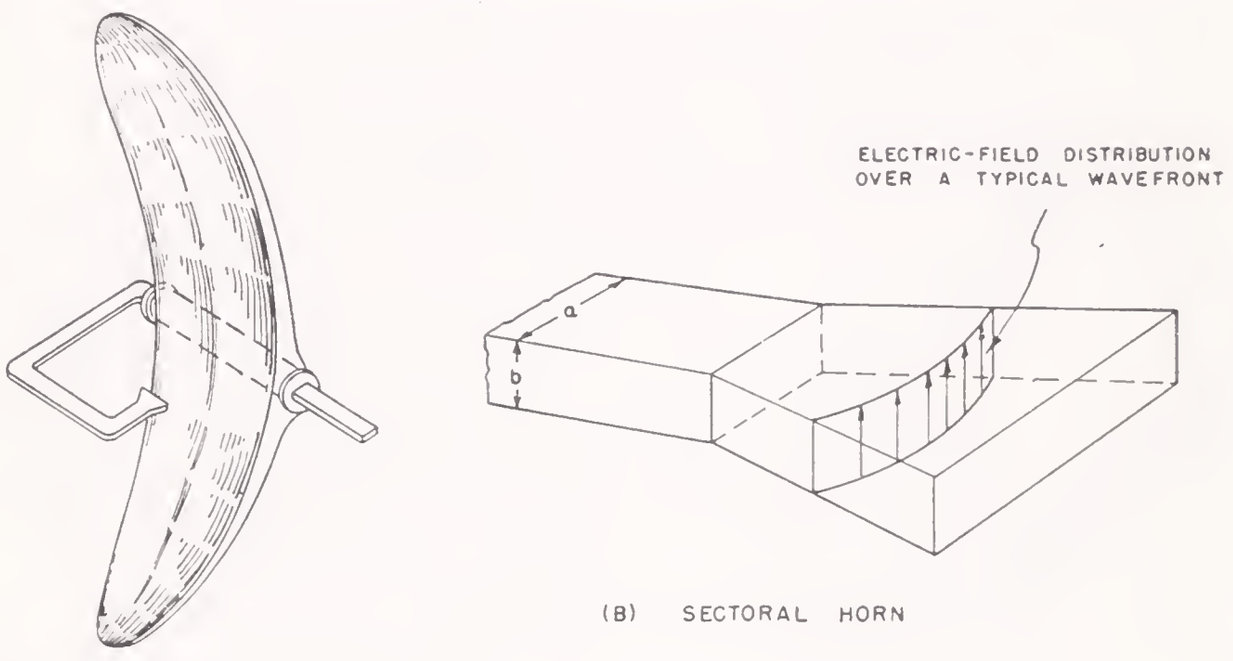


Fig. 47. Orange-peel paraboloid with sectoral horn feed.

same as the wavelength in space. Also, the wave impedance at the mouth of the horn approaches the value of wave impedance in space. Therefore, if the horn radiates into space, it provides almost the proper termination for the waveguide. In the presence of a parabolic reflector, satisfactory termination may often be had by adjustment of the horn dimensions and the spacing between the horn and the reflector. Sometimes additional impedance-changing devices are necessary.

Rear Waveguide Feeds. In Fig. 43*B* the half-wave antenna is so placed that nearly all the energy from the waveguide is absorbed and reradiated by the antenna. Thus the paraboloid is illuminated by a radiation pattern determined primarily by the antenna and the plane reflector which is one-quarter wavelength from the antenna.

A frequently used rear waveguide feed is the Cutler feed of Fig. 48*A*. The waveguide is terminated in a box of either circular or rectangular cross section. The termination may be considered as two small waveguides formed by dividing the main waveguide into two

halves and bending each half through 180 deg. The open ends, or slots, are a half wavelength apart and radiate toward the parabolic reflector. Figures 48*B* and *C* illustrate the construction of the termination and distribution of the field in the vicinity of the slots. Since the width of each slot is small compared with a wavelength, the primary radiation pattern in the electric plane is given by the $F^{2\frac{1}{2}}$ factor, except for the effect of the waveguide. The primary pattern in the magnetic plane depends upon the length of the slots and may be determined approximately from curve *B* in Fig. 42; for the usual slot lengths there

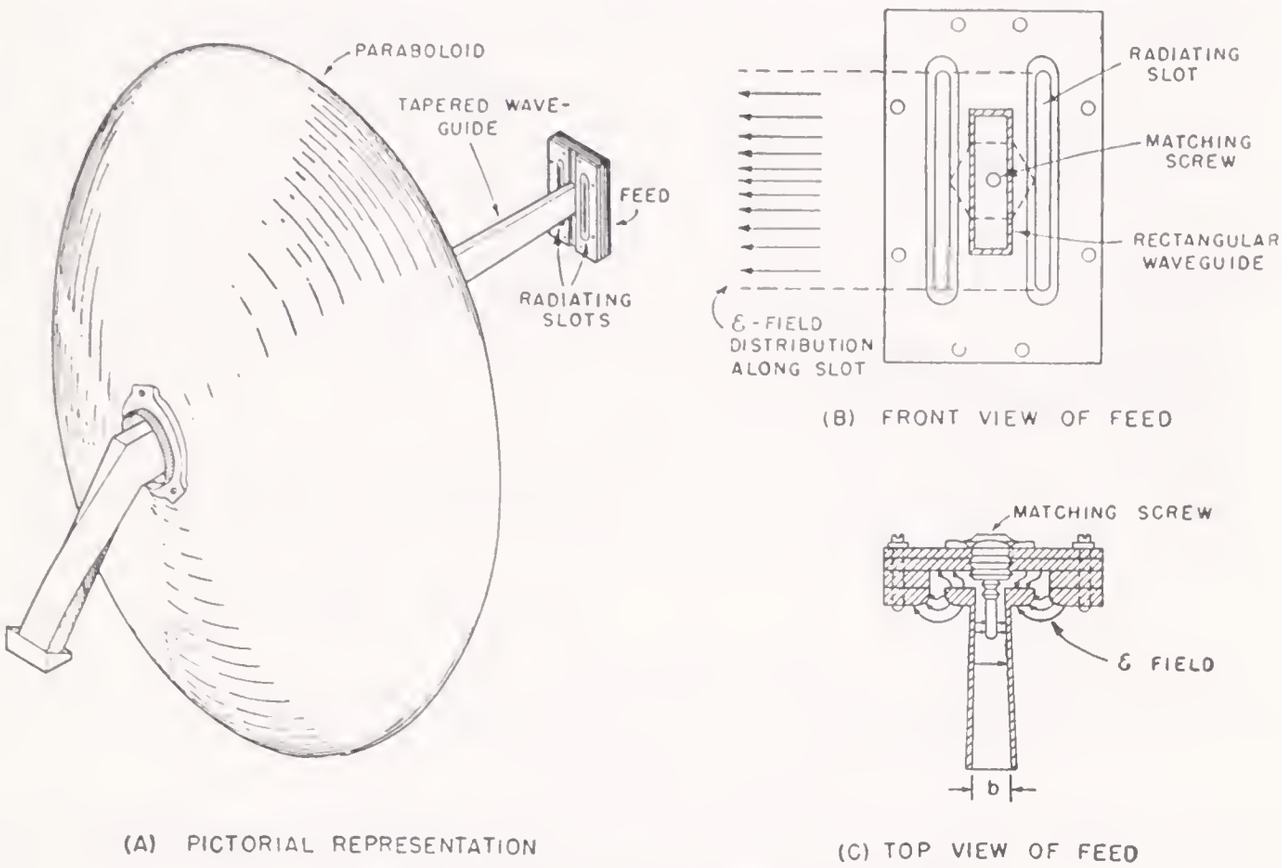


FIG. 48. Circular paraboloid with Cutler feed.

is not an important difference in the over-all beam angle of the antenna system in the two planes.

Because the slots are located a half-wavelength apart, the shorter cross-sectional dimension of the waveguide must be very small where the guide joins the termination. To connect this section of guide to the main waveguide, a tapered section is used so as to avoid the mismatch that would be caused by a sudden change in the *b* dimension. Tapering the guide in the *b* dimension does not affect the wave impedance of the guide; the actual matching of the terminating cavity to the guide is accomplished by adjusting the screw shown in Fig. 48*C*.

15. The Cosecant-squared Antenna. Because of the variety of the applications of radar, widely different requirements must be met by different antennas. Antennas designed to produce special polarization effects or pattern shapes may differ greatly from those already described.

An example of an antenna having a special pattern shape is the *cosecant-squared antenna* of Fig. 49. The top of the nearly paraboloidal reflector is bent forward to produce the special pattern. The name "cosecant-squared" arises because the power-density pattern varies (within physical limitations) as the square of the cosecant of the angle ϕ in Fig. 49*B*. Correspondingly, the field-intensity pattern varies as the cosecant of ϕ. The ideal shape of the field-intensity pattern is shown in Fig. 49*C*, together with the actual pattern obtained. The purpose of this pattern is to allow an air-borne radar system to lay

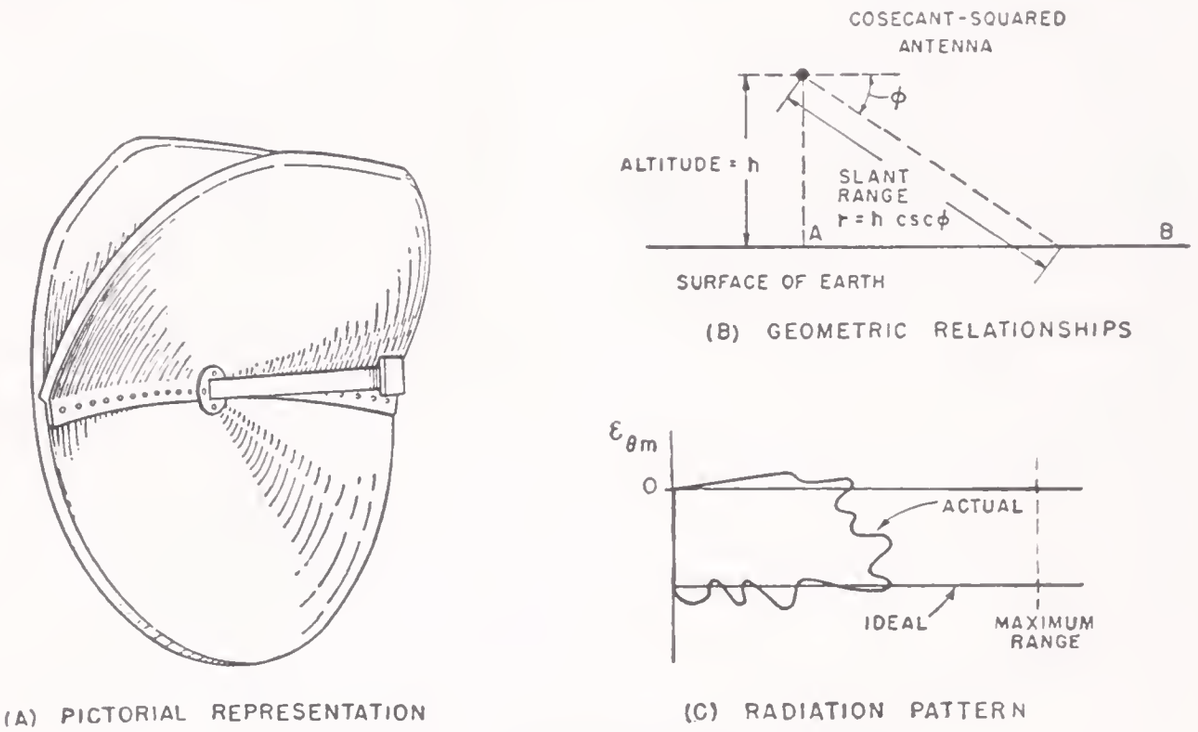


FIG. 49. The cosecant-squared antenna.

down a uniform electric field intensity along a line on the earth's surface such as line AB in Fig. 49*B*. The field intensity at a point on this line is given by

$$\mathcal{E}_{\theta m} = \frac{K}{r} \csc \phi$$

where K is a constant, r is the slant range to the point, and csc ϕ is a factor that describes the vertical-plane radiation pattern. Since r varies as csc ϕ (see Fig. 49*B*) the field intensity $\mathcal{E}_{\theta m}$ is a constant along the line AB. Because the same antenna is used for transmitting and receiving, equal indications are obtained on a range trace from similar targets at different ranges. This same type of antenna is also used (in an inverted position) on land-based and shipboard systems for obtaining equal indications from aircraft at equal altitudes but different ranges.

CHAPTER XIV

PROPAGATION

The performance of practically every type of radar system is modified to some degree by the effects of the surface of the earth on the radiation characteristics of the antenna and by the effects of both the surface and atmosphere of the earth on the propagation of radio waves between the radar set and the target. The influence of the earth and its atmosphere on radar-system performance is discussed briefly in this chapter.[1]

Land-based and shipboard equipments used for long-range search are most subject to the effects of the earth and its atmosphere because, at ranges greater than 100 miles, all targets appear at angles of elevation of 3 deg or less. The radiated energy must therefore travel close to the surface of the earth. On the other hand, gun-laying radar systems for heavy-caliber antiaircraft guns sometimes operate almost independently of the surface of the earth, particularly if the angle of elevation of high-altitude targets is larger than the radar beam angle when the targets come within gun range.

When radio waves are radiated at small angles of elevation, some of the waves strike the earth and are reflected. A portion of the reflected energy reaches the target along with waves which travel directly from the radar set to the target. As a result, the field intensity at the target differs from the field intensity in the absence of reflections, and the actual radiation pattern of the antenna differs from the free-space radiation pattern. The exact nature of these differences depends primarily upon the effective height of the center of the antenna above the reflecting surface and upon the roughness of surface and its conductivity and dielectric properties. The polarization of the radiated waves usually has a lesser effect.

Since high-frequency radio energy usually propagates along an approximately straight-line path, the curvature of the earth results in a horizon or line-of-sight limitation on maximum detection range. However, under certain weather conditions, abnormal refraction or bending of the path of propagation is caused by the atmosphere through which the energy travels, and the normal line-of-sight range is greatly exceeded. The

[1] For an extensive treatment of the subject of propagation at radar frequencies, see Donald E. Kerr, *Propagation of Short Radio Waves* (McGraw-Hill Book Company, Inc., New York, 1951).

effectiveness of land-based stations or shipboard equipment operating in the vicinity of land is influenced also by irregularities in the nearby terrain. A hill or mountain may act as a screen which prevents radar detection of targets beyond the obstruction.

1. Components of Radiation. To determine the radiation characteristics of an antenna located near the surface of the earth, it is convenient to divide the radiation into two parts: the *ground wave* and the *sky wave*, or *ionosphere wave*. The ground wave is further divided into a *direct wave*, *earth-reflected wave*, and a *surface wave*. Figure 1 shows the path of propagation of each of these components of radiation from a nondirectional antenna to a point P in space.

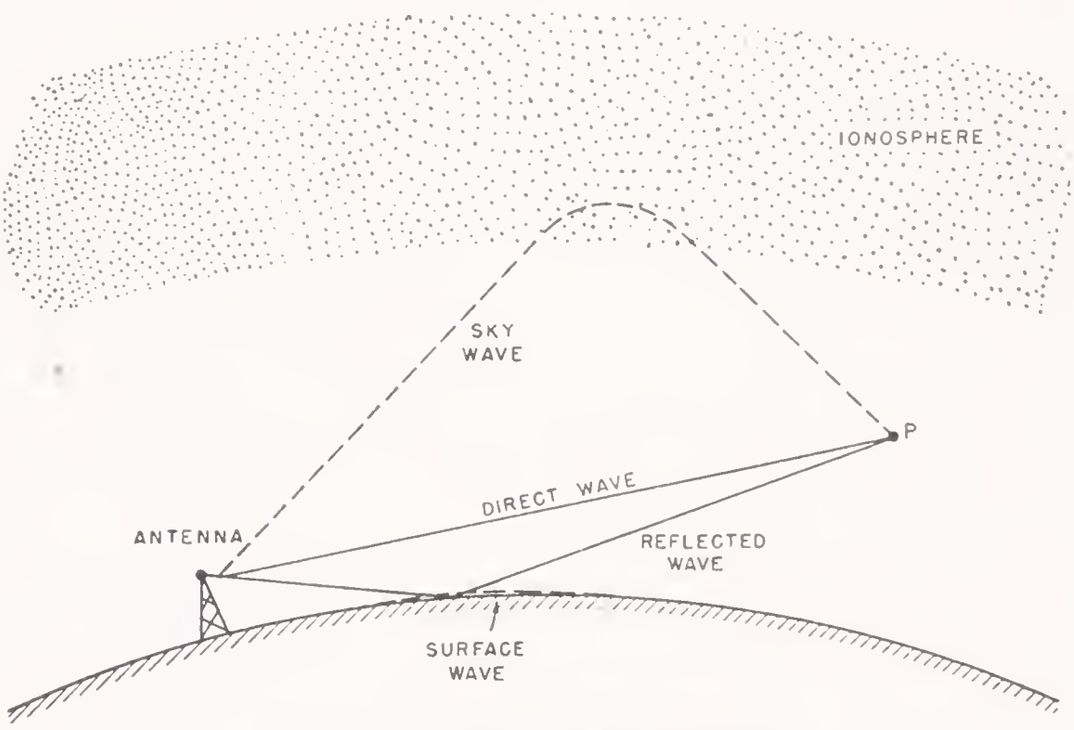


FIG. 1. Components of radiation.

The ground wave remains near the surface of the earth in traveling from the antenna to point P. The sky wave reaches point P after reflection from the ionosphere. Although important in many communication systems, the sky wave need not be considered in most radar applications because wavelengths shorter than about 4 m are not returned to the earth from the ionosphere. The components of the ground wave therefore make up the total signal reaching a target from a radar antenna.

The direct wave travels a nearly straight path from the antenna to point P. The reflected wave is reflected from the earth in such a way as to reach point P. Energy radiated from the nondirectional antenna of Fig. 1 strikes the earth at all points between the base of the antenna and the horizon, but only that ray which leaves the antenna in the direction shown reaches the point P.

The surface wave, which is the third component of the ground wave, is a wave diffracted around the surface of the earth or guided by the ground-

air interface. This component is important at broadcast frequencies, but at radar frequencies the surface wave is very rapidly attenuated, and at a distance of 1 mile from the antenna it has an amplitude of only a fraction of a per cent of the direct wave. This component must be considered in blind-landing systems in which ranges of less than a mile are important.

2. Reflection from a Perfectly Conducting Plane. Rigorous calculations for determining the radiation pattern of an antenna operating above the surface of the earth are complicated and laborious. Approximate results may be obtained by first considering the earth to be a perfectly conducting plane surface. Then the results of the simplified calculations

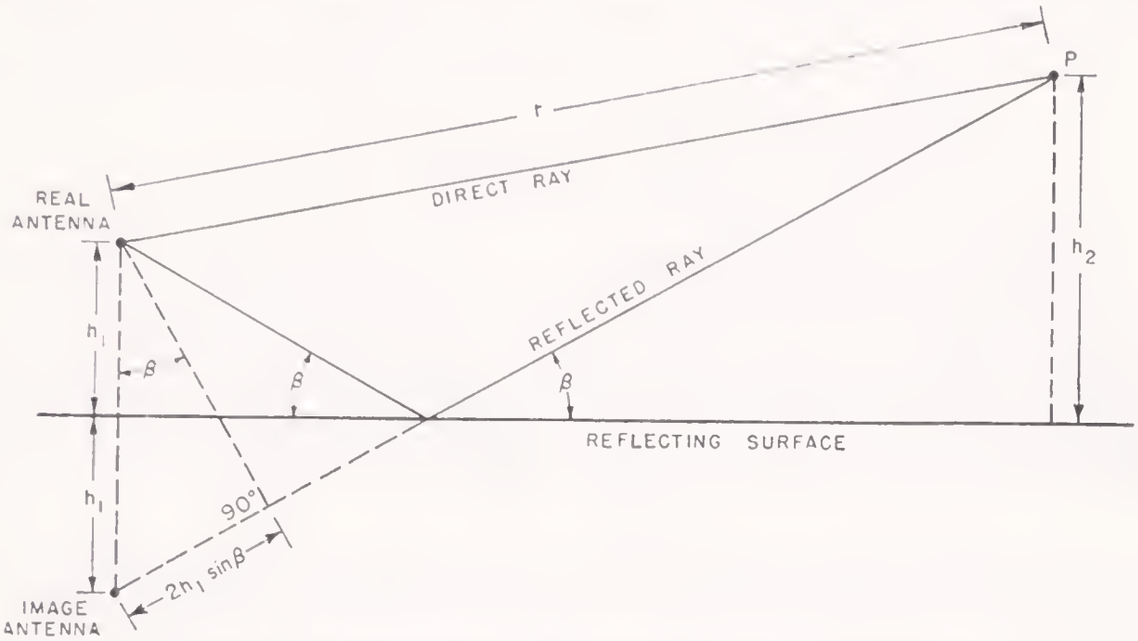


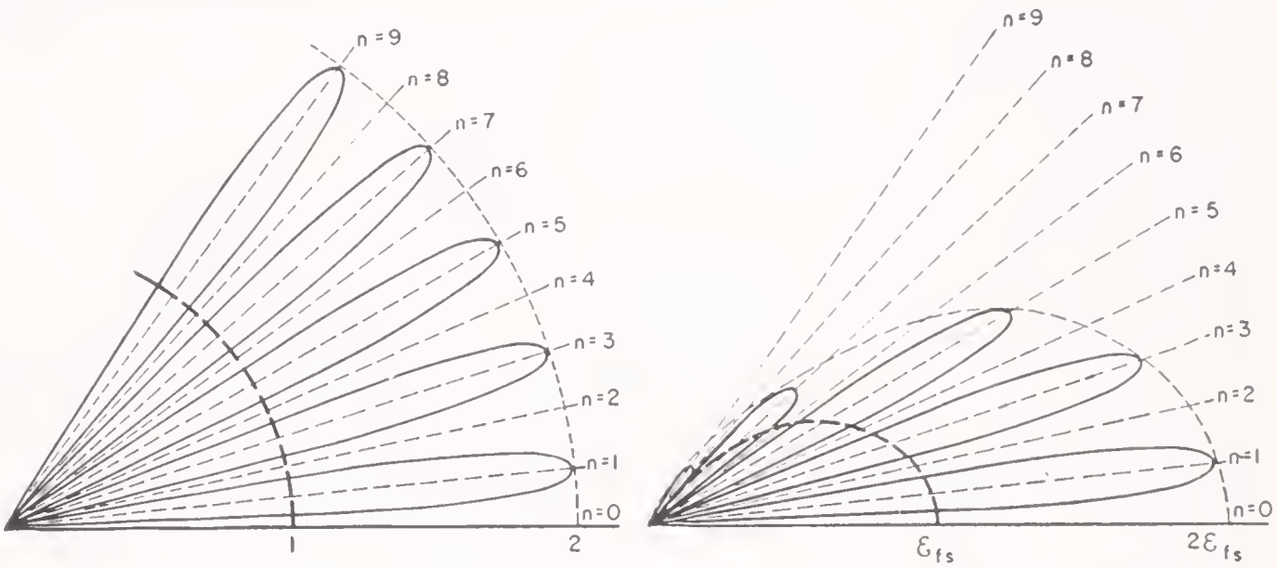
FIG. 2. Antenna above a perfectly conducting plane.

can be modified to include the effects of the finite conductivity and curvature of the earth, as well as atmospheric refraction.

Reflections from a perfectly conducting plane have been considered in the preceding chapter in connection with metal and screen reflectors for antenna arrays. The radiation pattern for an antenna and a reflecting screen is calculated by replacing the screen with an image antenna and by considering the real antenna and its image as a two-element array. The resultant pattern is represented by a G factor if the real antenna is parallel to the screen, since the image must be energized 180 deg out of phase with the original antenna to take into account the phase reversal at the reflecting screen. If the real antenna is perpendicular to the screen, no phase reversal occurs at the reflecting surface; hence the equivalent image is in phase with the real antenna, and the resulting pattern is represented by an F factor. Use of the F and G factors to obtain the radiation pattern of an antenna in the presence of a perfectly conducting plane earth is the same as the use of these factors to obtain the radiation pattern of an array with a reflecting screen, except that the spacing

between driven elements and the reflecting screen of an antenna array is usually a quarter wavelength or less while the spacing between an antenna array and the earth is usually many wavelengths.

Let the peak value of the free-space, vertical-plane radiation pattern of the real antenna in Fig. 2 be represented by $\mathcal{E}_{fs} = (60I_m/r)F_a$, where F_a is the product of all F and G factors that contribute to the formation of the free-space pattern and I_m is the peak current in each element of the array. Then, in the presence of the earth, the pattern includes another factor which accounts for the contribution of the reflected ray or its equivalent, the image antenna.



(A) G FACTOR INTRODUCED BY REFLECTION (B) VERTICAL-PLANE RADIATION PATTERN

FIG. 3. Reflection of horizontally polarized waves from a perfectly conducting plane.

For a horizontally polarized antenna the complete vertical-plane pattern is represented by $\mathcal{E}_{\theta m} = (60I_m/r)F_aG_{\frac{1}{2}}{}^{2(2h_1/\lambda)}$. For a vertically polarized real antenna the complete vertical-plane pattern is

$$\mathcal{E}_{\theta m} = (60I_m/r)F_aF^{2(2h_1/\lambda)}$$

The free-space pattern usually has a single major lobe and secondary lobes of negligible significance. For large values of h_1/λ, both the F and G factors for earth reflection have a large number of nulls and therefore a large number of very narrow lobes, all of which have the same amplitude. A typical G factor is plotted in Fig. 3*A*. The product of the free-space pattern and the reflection factor is a pattern which has the same general outline as the free-space pattern but with the major lobe broken up into a large number of very sharp lobes, as indicated in Fig. 3*B*.

The G factor for horizontal polarization always produces a null in the field intensity along the surface of the reflector. The F factor for vertical polarization produces a maximum field intensity at the surface of the reflector. Accordingly, vertical polarization would seem to be superior to horizontal polarization for detection of surface targets and low-flying air-

craft. However, when the curvature and finite conductivity of the earth are taken into account, this apparent advantage disappears (see Art. 3 and 5).

The horizontal-plane radiation pattern in the presence of the earth is the same as the free-space pattern and is not considered in this discussion.

The calculations necessary for a complete tabulation of the F and G factors for large values of h_1/λ are tedious because of the numerous lobes in the patterns. To obtain approximate results by graphical multiplication of patterns, it is sufficient to determine the angles of elevation at which maxima and minima occur. As shown in Fig. 2, the difference in path length of the direct and reflected (or image) rays is $2h_1 \sin \beta$. For *horizontal polarization* the real and image antennas are 180 deg out of phase; hence, if the path difference is

$$2h_1 \sin \beta = \frac{n\lambda}{2} \tag{1}$$

there are maxima in the pattern for all odd-integral values of n and minima for even-integral values of n. For vertical polarization the conditions for maxima and minima are reversed. To determine the angles at which maxima and minima occur, first solve Eq. 1 for sin β and obtain

$$\sin \beta = \frac{n\lambda}{4h_1} \tag{2}$$

At long ranges the elevation angles of aircraft targets are only a few degrees, and since the sine of a small angle is approximately equal to the angle in radians,

$$\beta = \frac{n\lambda}{4h_1} \qquad \text{radians} \tag{3}$$

or

$$\beta = 57.3 \frac{n\lambda}{4h_1} = 14.3 \frac{n\lambda}{h_1} \qquad \text{deg} \tag{4}$$

The factor for reflections produces many maxima and minima within a small vertical angle. For example, if $h_1 = 10\lambda$, Eq. 4 shows that the first maximum obtained with horizontal polarization occurs at an angle of elevation of 1.43 deg. As h_1/λ increases, the angle at which the first maximum occurs decreases and the number of maxima and minima within a given vertical angle increases. If the complete pattern is desired, it can be sketched by recalling that the beam angle of an antenna lobe measured between half-power points is approximately half of the angle between nulls.

The patterns in Fig. 3 are drawn for $h_1/\lambda = 2.4$, which fixes the angle of elevation of the lowest lobe at about 6 deg. All the lobes of the G

factor introduced by reflections have a maximum value equal to 2. The maximum values of the lobes in the antenna radiation pattern follow a contour similar to the free-space pattern itself, and the lowest lobe has a maximum value nearly twice as large as the maximum value of the main lobe in the free-space pattern. In practice antenna heights are much greater, and as a result, there are many more lobes in the pattern and the lowest lobes occur at much smaller angles. At the very small angles of elevation which are important for long-range search radar, the free-space pattern of the antenna has a negligible effect on the shape of the resultant vertical-plane pattern.

3. Reflection Coefficients. Reflection from either a perfectly conducting surface or from the earth can be expressed in terms of a reflection coefficient Γ which is the ratio of the reflected field intensity to the incident field intensity. This quantity is analogous to the reflection coefficient associated with traveling waves on transmission lines.

For a perfectly conducting surface, there is no absorption loss in the reflecting surface; hence the magnitude of the reflected field intensity is equal to the magnitude of the incident field intensity, and the reflection coefficient has a magnitude of one. For horizontal polarization there is a phase reversal at the surface, so that $\Gamma_h = |\Gamma_h|\epsilon^{j\phi_h} = \epsilon^{j180^\circ}$. For vertical polarization there is no phase reversal at the reflecting surface; hence $\Gamma_v = |\Gamma_v|\epsilon^{j\phi_v} = \epsilon^{j0^\circ}$.

For many long-range-search radar installations, the reflecting surface is the sea; hence the reflection coefficient of sea water at radar frequencies is important. This coefficient has been calculated and checked experimentally, and calculated and experimental results have been found to agree quite closely provided the sea is smooth. For a sea of appreciable roughness, reflections are likely to be very irregular and to fluctuate rapidly with time, particularly at extremely short wavelengths. The effect of roughness is also more noticeable with horizontal polarization than with vertical polarization.

Figure 4 shows the variation of magnitude and phase angle of the reflection coefficient $\Gamma = |\Gamma|\epsilon^{j\phi}$ of a smooth sea with changes in angle of incidence β for both horizontal and vertical polarization. Observe that, for horizontally polarized radio waves in the frequency range from 100 to 3,000 Mcps, the reflection coefficient of smooth sea water is very nearly the same as the reflection coefficient for a perfect conductor. For vertical polarization there is a significant difference between the reflection coefficient for sea water and for a perfect conductor. In fact, sea water acts more nearly like a dielectric. At the angle for which $|\Gamma_v|$ is a minimum, there is maximum transmission of energy into the surface.

Figure 5 compares the reflection patterns obtained for horizontal and vertical polarizations, assuming a smooth-plane reflecting surface of sea

water, a frequency of 3,000 Mcps, an antenna height $h_1 = 14.3\lambda$, and a nondirectional antenna. For convenience in plotting, angles have been multiplied by a factor of 5, so that 5 deg on the figure represents 1 deg in space.

There is a null along the surface for both horizontal and vertical polarization, since ϕ_v as well as ϕ_h is 180 deg at grazing angle of incidence. As the angle of incidence is increased, the angles at which maxima and minima occur in the vertical polarization pattern deviate from those for horizontal polarization, since ϕ_v deviates from 180 deg.

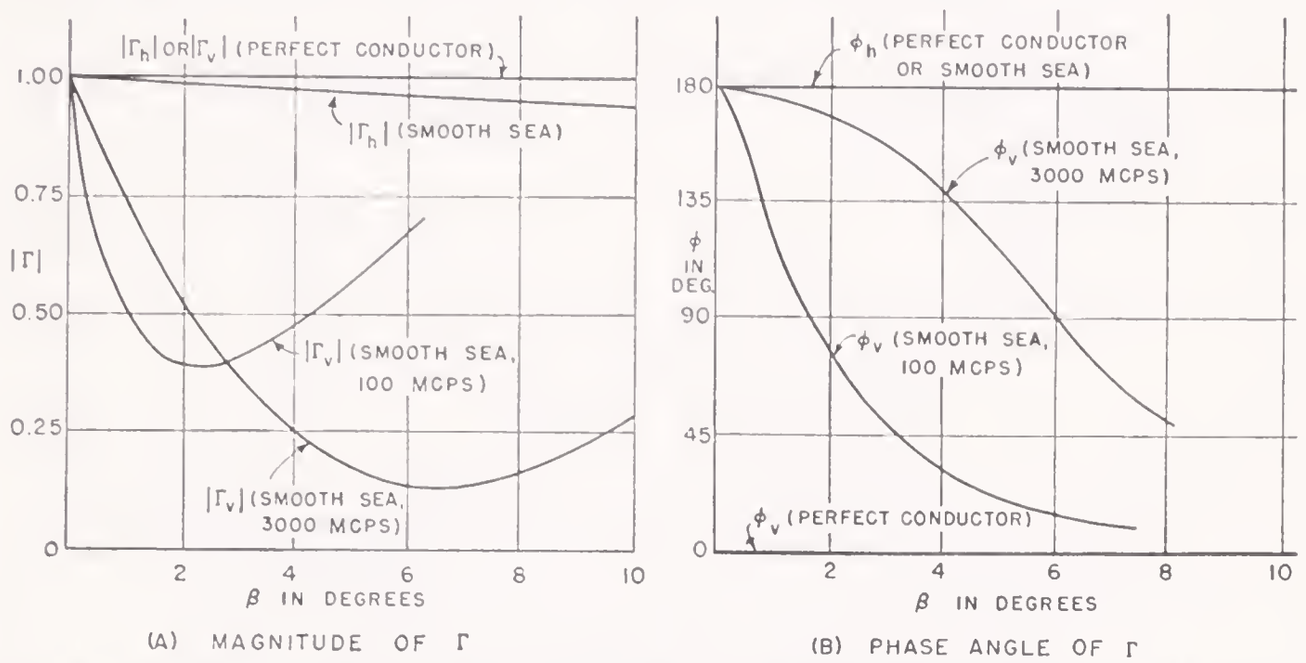


FIG. 4. Variation of magnitude and phase angle of reflection coefficient $|\Gamma|\epsilon^{i\Phi}$ of a smooth sea for angles of incidence β from 0 to 10 deg. [*Adapted from "Relation of Radar Range to Frequency and Polarization," Radiation Laboratory Report C-6 (M.I.T., Cambridge, Mass., 1942), with permission.*]

For horizontal polarization, the electric-field intensity at the maxima of the lobes is nearly twice the free-space field intensity, because $|\Gamma_h| \approx 1.0$. For vertical polarization the deviation from the free-space field intensity is small at relatively large angles of incidence, because $|\Gamma_v|$ is small. For example, the lobes in the vicinity of 5 deg have maximum amplitudes of 1.98 and 1.2 for horizontal and vertical polarization, respectively.

4. Effect of Earth Reflections on Range. The maximum range at which a given target can be detected by a radar set is greater in the presence of sea reflections than it is in free space, provided the target is located at a point where the direct and reflected waves are in phase. A radar set operating in free space produces a field intensity at the target equal to k'/r, where r is the range to the target. The reflection from the target produces a field intensity at the radar receiver equal to k/r^2, where k is a new constant of proportionality which includes k' (see Art. 5, Chap. I). Maximum range r_{max} corresponds to the minimum detectable signal k/r_{max}^2.

In the presence of sea reflections and with the same target at free-space maximum range, the field intensity at the target is nearly $2k'/r_{max}$ for horizontal polarization, provided the target is located at a point where direct and reflected waves are in phase. The echo signal returning to the radar receiver comprises a direct wave and a reflected wave, each of which produces a field intensity at the radar receiver of $2k/r_{max}^2$. The direct and reflected waves arrive at the receiver in phase, because they follow the identical paths taken by the transmitted signal. Therefore, the total field intensity at the receiver is $4k/r_{max}^2$, which is four times the minimum detectable field intensity. The maximum range at which a target can be

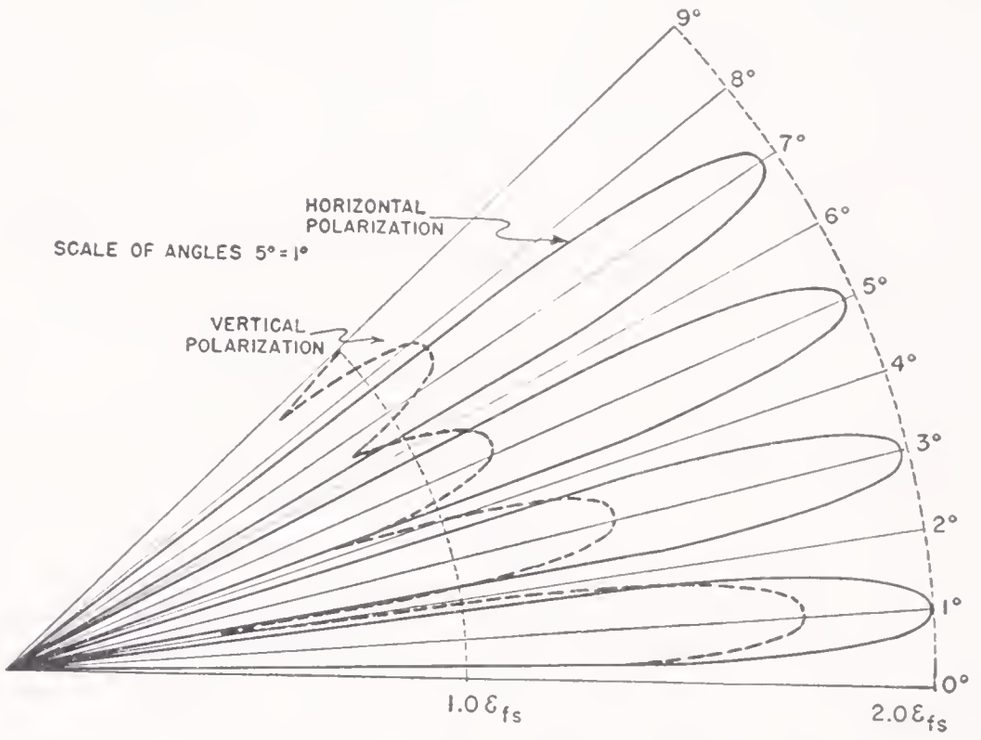


FIG. 5. Effect of polarization on the vertical-plane radiation pattern of a nondirectional 3,000-Mcps antenna located 14.3λ above a smooth-plane reflecting surface of sea water.

detected is twice the maximum free-space range, because at a range $2r_{max}$ the field intensity returned to the receiver is $4k/4r_{max}^2 = k/r_{max}^2$. Thus the presence of sea reflections approximately doubles maximum detection range and hence is equivalent to a sixteenfold or 12-db increase in transmitted power.

At points where direct and reflected waves arrive at the target out of phase, the detection range is less than in free space; hence there are gaps or blind spots between the lobes of the vertical-plane pattern.

Since maximum range is proportional to field intensity at the target, it is evident from Fig. 5 that, for a smooth surface, horizontal polarization yields greater maximum ranges than vertical polarization but that the gaps between lobes are more pronounced. However, the advantage in range of detection outweighs the disadvantage of the gaps. For most microwave systems, h_1/λ is large and the lobes are so closely

spaced that the pattern can be considered continuous. At frequencies of 100 or 200 Mcps, the lobes are ordinarily widely spaced. Two antenna systems may then be used having values of h_1/λ so chosen that the lobes of one occur at the gaps of the other.

5. Divergence. The results obtained for reflection from a plane surface indicate, in a general way, the effect of a reflecting earth on the vertical-plane radiation pattern of an antenna. These results are modified by the curvature of the earth as a result of the divergence, or spreading, of the rays when they are reflected from a curved surface. Because divergence is the result of the spherical shape of the earth,

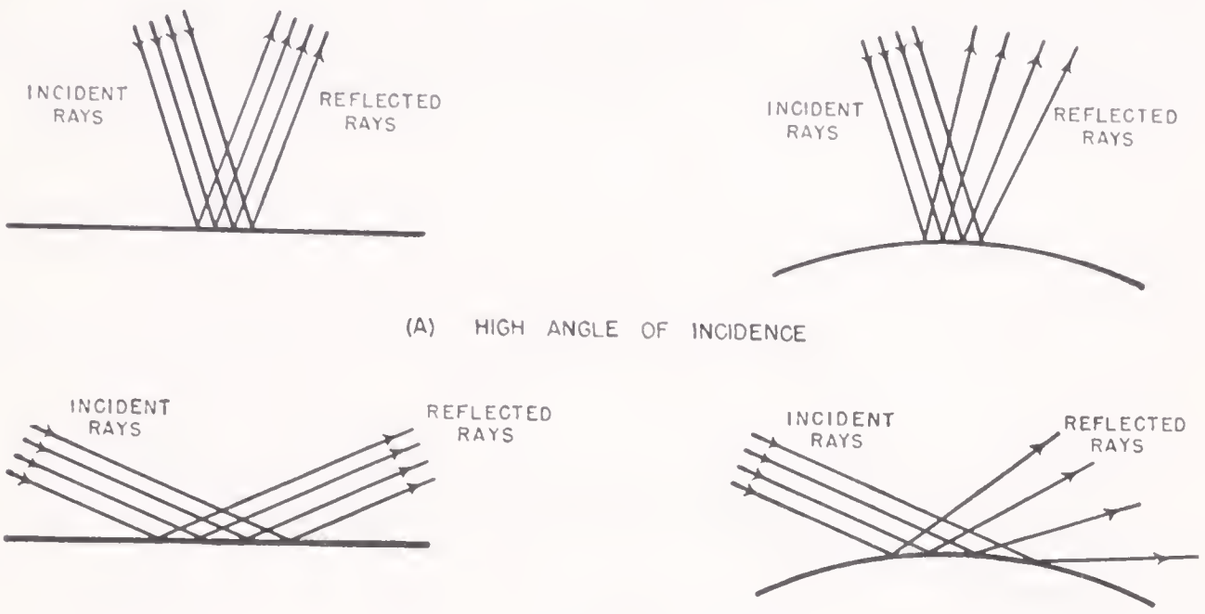


FIG. 6. Divergence caused by a curved surface.

it affects horizontally and vertically polarized waves equally. The effect is greatest for grazing angles of incidence and least for rays incident normal to the surface.

The divergence of rays after reflection from a curved surface is illustrated in Fig. 6, where the effects of a plane surface and of a curved surface on a bundle of rays in a plane radio wave are compared for high and low angles of incidence. The beam incident on the plane surface has the same power density as the beam incident on the curved surface.

For the plane surface the reflected beam has the same cross-sectional area as the incident beam, since angles of incidence and reflection are equal. Also, the reflected power density is the same as the incident power density. For the curved surface, the cross section of the reflected beam is larger than that of the incident beam. Thus the power density is smaller after reflection. The divergence is greatest for grazing angles of incidence because a beam of given cross section impinges on a greater area of the curved surface. In the calculations for vertical-plane radiation patterns over a curved surface such as the sea, the effects of

divergence are included by a divergence factor[1] which varies between zero at grazing incidence and unity at normal incidence. The zero value occurs at grazing incidence because, as shown in Fig. 7, all reflected waves strike the earth between the radar set and the radar horizon and none reach the target. Within a very small angle above the line of sight, the value of the divergence factor rises to about 0.5 and then rises slowly toward 1.0 as normal angles of incidence are approached. At very small angles above the line of sight, the divergence factor controls the magnitudes of the lobes in the vertical-plane radiation pattern, whereas at large angles, the free-space pattern of the antenna determines the size of the lobes.

6. Standard Atmospheric Refraction. In the preceding articles, the medium through which the radio waves are propagated is assumed to be

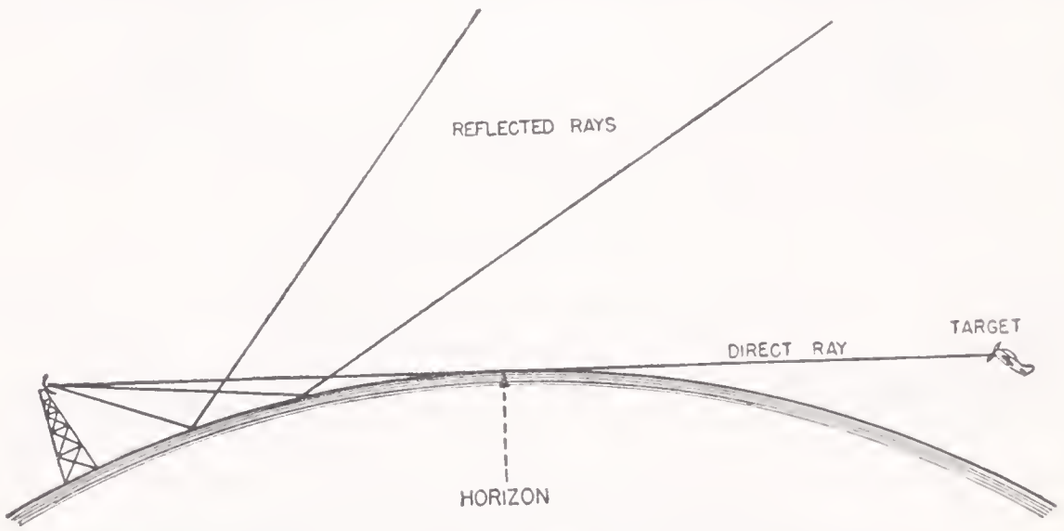


FIG. 7. Target on radar line of sight.

uniform; that is, the medium is considered to have a constant index of refraction or dielectric constant so that velocity of propagation is constant and the rays are straight lines. In the atmosphere of the earth, the temperature, partial pressure of the dry air, and the water-vapor content normally decrease with increasing altitude, and therefore the index of refraction decreases with altitude. The velocity of propagation of electromagnetic waves varies inversely with the index of refraction of the medium; hence the radio waves move slightly more rapidly in the upper atmosphere than they do close to the surface of the earth, and the result is a downward bending of the rays or paths of propagation, as shown in Fig. 8.

In the presence of standard refraction, the curvature of the rays is less than the curvature of the earth, and a *radar horizon* exists beyond the optical horizon, as indicated in Fig. 8. Standard refraction is assumed to occur when the index of refraction decreases uniformly at the rate of

[1] See F. E. TERMAN, *Radio Engineers' Handbook*, 691, 692 (McGraw-Hill Book Company, Inc., New York, 1943).

1.2×10^{-8} per foot increase in altitude. Because of the linear decrease of the index of refraction with increase in altitude, the velocity of propagation increases linearly with altitude, and the ray paths are very nearly arcs of circles over the distances involved in radar detection. Since radiation patterns or coverage diagrams (see Art. 8) are more conveniently plotted for straight-line propagation, it is desirable to assume

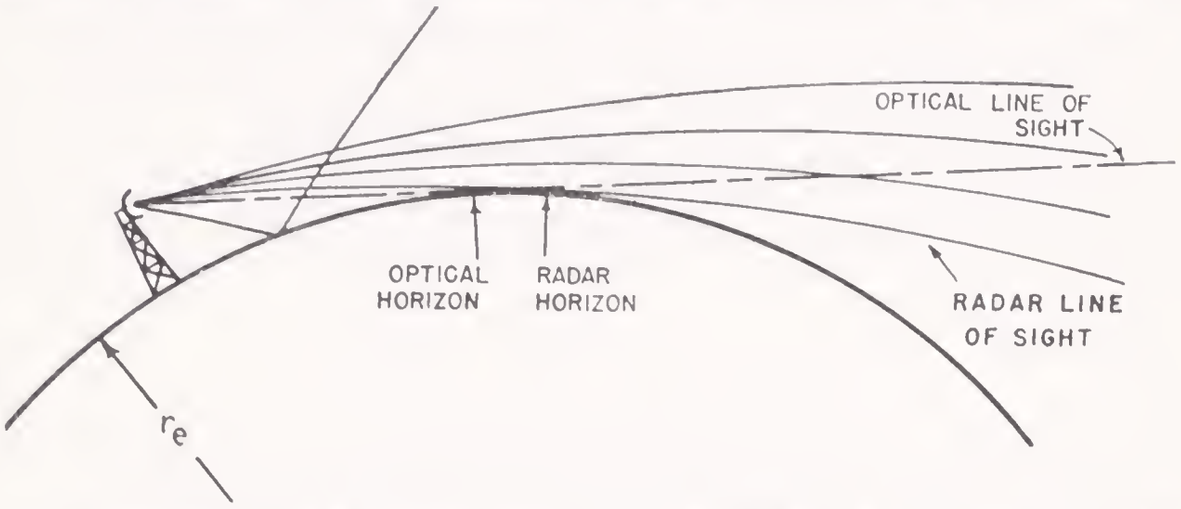


FIG. 8. Extension of radar horizon beyond optical horizon as the result of standard atmospheric refraction.

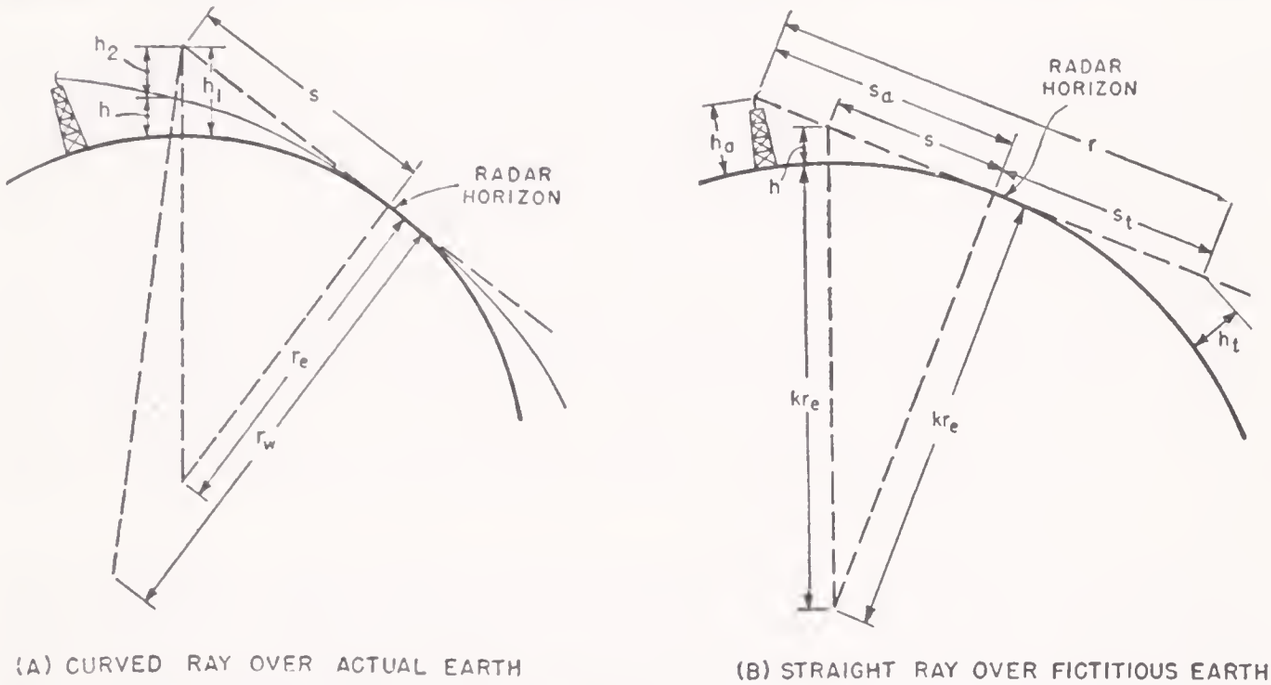


FIG. 9. Relationship between curved path of ray over actual earth and straight path of ray over earth of increased radius.

a fictitious earth having a larger radius than that of the actual earth, so that a straight ray from the antenna is tangent to the fictitious earth at the radar horizon, as shown in Fig. 9*B*.

For equivalence of results obtained with the fictitious earth of Fig. 9*B* and the curved path and actual earth of Fig. 9*A*, the height h of the ray above the earth must be the same in both figures for a given value of s. From the right triangle in Fig. 9*B*, the following relationship is obtained:

$$s^2 + (kr_e)^2 = (h + kr_e)^2 = h^2 + 2hkr_e + (kr_e)^2 \tag{5}$$

In Eq. 5, h^2 is negligible relative to $2hkr_e$, and therefore the equation reduces to

$$s^2 = 2hr_e \tag{6}$$

Thus

$$h = \frac{1}{2}\left(\frac{1}{kr_e}\right)s^2 \tag{7}$$

In a similar way the right triangles in Fig. 9*A* yield the relations:

$$h_1 = \frac{1}{2}\left(\frac{1}{r_e}\right)s^2 \qquad \text{and} \qquad h_2 = \frac{1}{2}\left(\frac{1}{r_w}\right)s^2 \tag{8}$$

where r_w is the radius of the curved path of the wave. Therefore in Fig. 9*A*,

$$h = h_1 - h_2 = \frac{1}{2}\left(\frac{1}{r_e} - \frac{1}{r_w}\right)s^2 \tag{9}$$

Comparison of Eqs. 7 and 9 shows that Figs. 9*A* and *B* are equivalent, provided k is chosen to make

$$\frac{1}{kr_e} = \frac{1}{r_e} - \frac{1}{r_w} \tag{10}$$

that is, to make the curvature of the fictitious earth equal the difference between the curvatures of the actual earth and of the path of the ray.

The quantity kr_e is called the *effective radius of the earth*. For standard refraction $r_w = 4r_e$ and hence $k = \frac{4}{3}$. For nonstandard conditions the value of k may vary from 1.1 to 1.6 or more. The smaller values are more likely to exist in cold, dry climates or at great heights, whereas the larger values occur more frequently in tropical, humid climates. Local weather conditions may produce much greater variations in the value of k. Figure 10 is obtained when the rays in Fig. 8 are replotted as straight lines over an earth of radius $\frac{4}{3}r_e$.

Standard refraction is usually assumed in determining approximately the limiting detection range imposed by the radar horizon. If the height of the radar antenna is h_a and the altitude of the target is h_t, as indicated in Fig. 9*B*, Eq. 6 may be used to show that the limiting range is

$$r = s_a + s_t = \sqrt{2h_akr_e} + \sqrt{2h_tkr_e} \tag{11}$$

To make this equation more convenient to use, r may be expressed in miles and h in feet, and numerical values may be inserted for k and r_e. The mean radius of the earth is 3,960 miles, and when $k = \frac{4}{3}$, $kr_e = 5{,}280$ miles. Since there are 5,280 ft in a mile, the numerical constants cancel, and

$$r = \sqrt{2h_a} + \sqrt{2h_t} \qquad \text{miles} \tag{12}$$

where h_a and h_t are in feet.

7. Vertical-coverage Diagrams. Comparison of detection ranges is facilitated by converting vertical-plane radiation patterns to vertical-coverage diagrams. A vertical-plane radiation pattern usually represents a plot of measurements or calculations of field intensity versus a variable angle for a fixed distance between transmitter and receiver. Mathematically, the radiation pattern is a plot of $\mathcal{E}_{\theta m} = (k_1/r)F(\beta)$ for a fixed value of r. The factor $F(\beta)$ includes all the variable factors for the free-space pattern as well as the factor for sea reflection, and k_1 is the amplitude constant. If this equation is solved for r and $\mathcal{E}_{\theta m}$ is held constant, the result is $r = (k_1/\mathcal{E}_{\theta m})F(\beta)$. This equation represents a constant-field-intensity contour and has exactly the same form as the radiation pattern, since the variable factor $F(\beta)$ is the same.

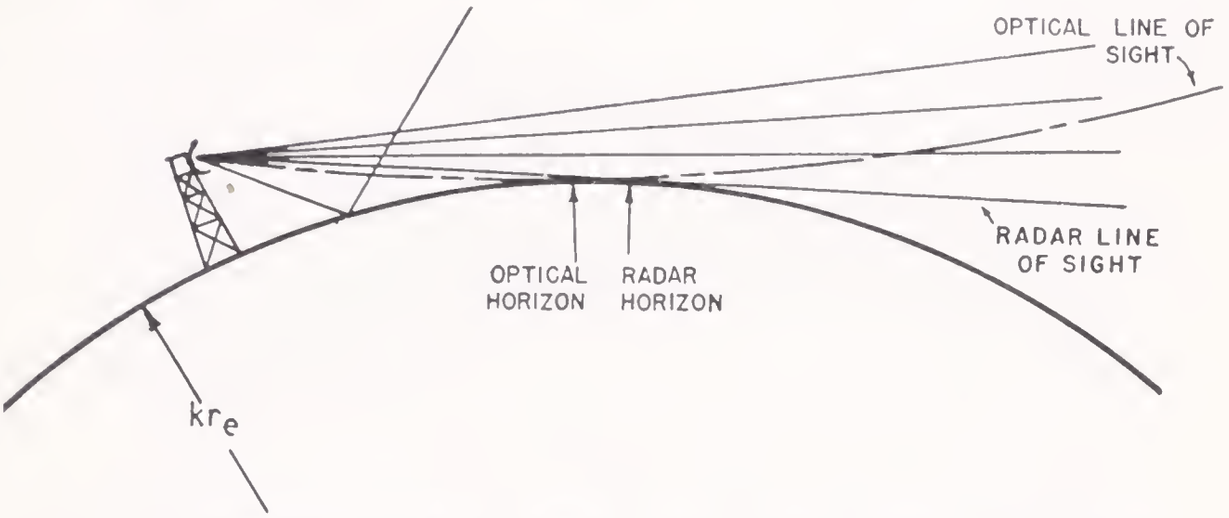


FIG. 10. Straight-line radar propagation over a fictitious earth of radius $\frac{4}{3}r_e$.

For a radar set which employs a single antenna system, the received field intensity $\mathcal{E}_{rm}$ is

$$\mathcal{E}_{rm} = \frac{k_1}{r} F(\beta) \times \frac{k_2}{r} F(\beta) = \frac{k_1 k_2}{r^2} [F(\beta)]^2 \tag{13}$$

since the directivity of the radiation system, which includes the antenna and the earth, affects both the outgoing and the echo signals. If Eq. 13 is solved for r, the result is

$$r = \sqrt{\frac{k_1 k_2}{\mathcal{E}_{rm}}} F(\beta) \tag{14}$$

and if this equation is used to plot r as a function of β for $\mathcal{E}_{rm}$ equal to the minimum detectable field intensity, the resulting curve is a *vertical-coverage diagram*. Any target inside the lobes of this pattern is detectable by the radar set, but targets outside of the lobes are not detectable. Note that the vertical-coverage diagram has the shape of the factor $F(\beta)$, which is the same as the shape of the one-way radiation pattern or the constant-field-intensity contour.

To convert a radiation pattern to a vertical-coverage diagram, it is only necessary to adjust the scale of the plot in order to make the calculated pattern fit the performance of the radar set. This correlation is usually accomplished by making observations on the radar indicator of a target plane flying specified courses. This technique, called calibration, is particularly important in checking uhf radar systems for blind spots or gaps in the coverage patterns. The calibration procedure also shows up the effects of irregularities in the terrain or of reflecting objects near the radar antenna or of any other factors which cannot be included in the calculation of vertical-coverage diagrams. Vertical-coverage diagrams are usually drawn to a mixed scale with range measured in miles and

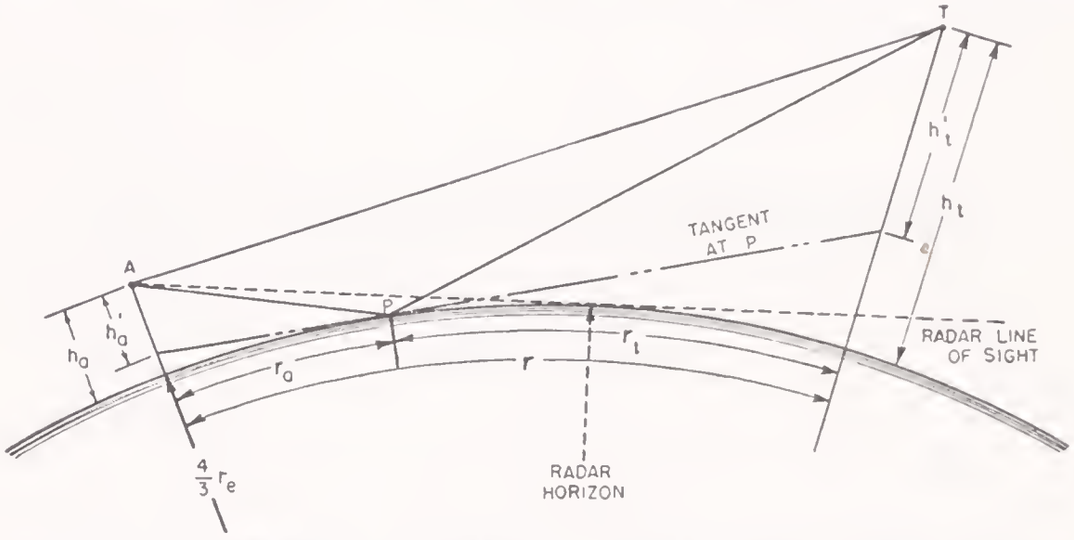


FIG. 11. Diagram for calculation of coverage over a curved earth.

altitude measured in feet (see Fig. 12). With some of the early types of radar sets, the coverage diagrams are used to determine the approximate altitude of a target by noting the range of first detection and the ranges of successive fade and redetection points as the target passes through the lobes of the pattern.

The curved-earth geometry involved in the calculation of the form of the vertical-coverage diagram is shown in Fig. 11. Standard refraction is assumed; hence the radius of the earth is multiplied by the factor $\frac{4}{3}$. The radar line of sight is drawn, and the relative field intensity is calculated for each target position above the line of sight. The calculations for any one point involve several steps. The reflection point P must be located so that angles of incidence and reflection are equal for the reflected ray. A tangent line is drawn at P, and from the heights h_a' and h_t', the difference in path lengths for direct and reflected rays may be determined. From this difference and from a knowledge of the reflection coefficient and divergence factor for the angle of incidence involved, the total field at the target is obtained. Although the method as outlined is relatively simple, the calculations are laborious. Charts and formulas for use in making coverage diagrams have been published.

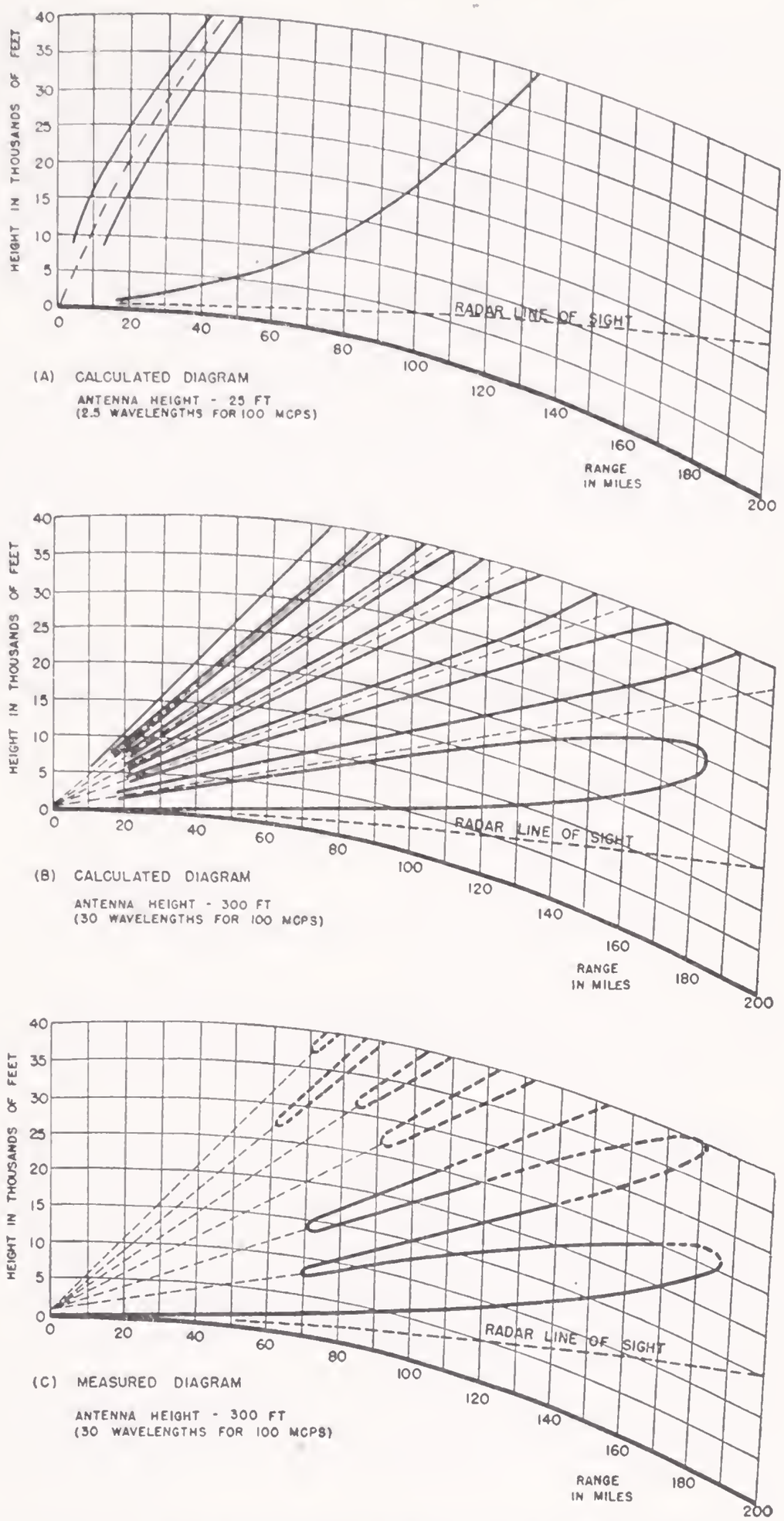


FIG. 12. Vertical-coverage diagrams, calculated and measured.

Typical radar-coverage diagrams are shown in Fig. 12. These are drawn for standard refraction and therefore do not give the true coverage under nonstandard conditions. Also, they do not include the decrease in detection range caused by interfering signals, such as sea return or ground clutter. The irregularities in the reflecting terrain which cause sea return or ground clutter also introduce multiple reflection points which tend to make the lobe structure less pronounced. This effect is equivalent to a decrease in the effective magnitude of the reflection coefficient and is more pronounced for microwave radar than for uhf radar.

8. Nonstandard Refraction. Vertical-coverage diagrams serve only as a basis for estimating the average performance to be expected of a radar set. Variations in atmospheric conditions affect the amount of refraction of the radar waves and modify the radar coverage. When atmospheric refraction is greater than standard, the rays more nearly

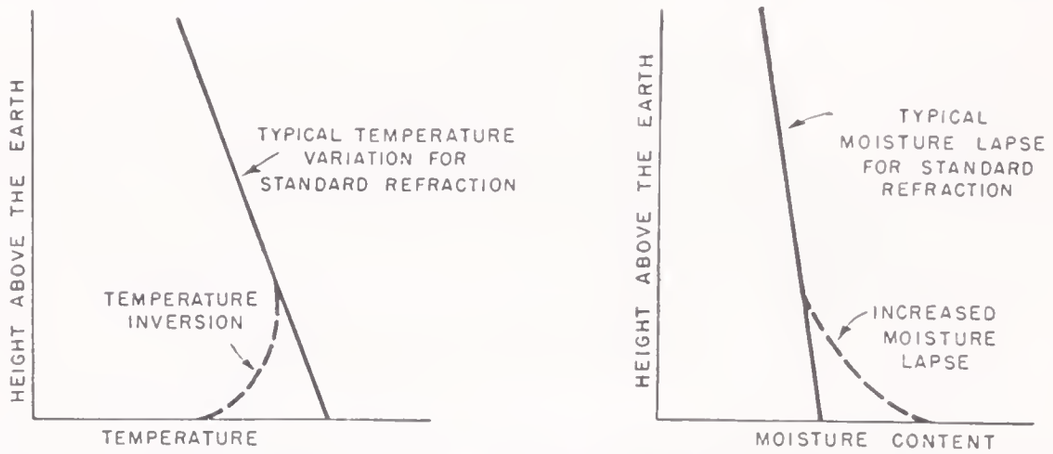


FIG. 13. Profiles of temperature and moisture content of the atmosphere.

follow the earth and detection ranges on low-flying or surface targets are markedly increased. When refraction is less than standard, the rays follow more nearly straight lines and maximum detection ranges are reduced.

Under conditions of standard refraction, the temperature and moisture content of the air decrease relatively slowly from the surface of the earth to altitudes of several thousand feet. The straight-line profiles in Fig. 13 represent typical variations of temperature and moisture content of the atmosphere with height above the earth for standard refraction. One of the conditions which tends to bring about abnormal refraction results from the passage of warm, dry air over a cool body of water. Evaporation of water from the surface produces a concentration of moisture and a decrease in temperature near the surface. This increased moisture lapse[1] and the accompanying temperature inversion are shown in Fig. 13 by

[1] The lapse of a quantity x is defined as $-dx/dh$, where h denotes height. Thus an increase in moisture lapse means a more rapid decrease in moisture content with height.

dashed lines. When these conditions exist, the atmospheric refraction is increased.

The temperature-moisture variations shown in Fig. 13 may extend from the surface to a height of only a few feet or to a height of a few hundred feet. When the temperature inversion and moisture lapse are very pronounced, the curvature of the rays may be equal to, or even greater than,

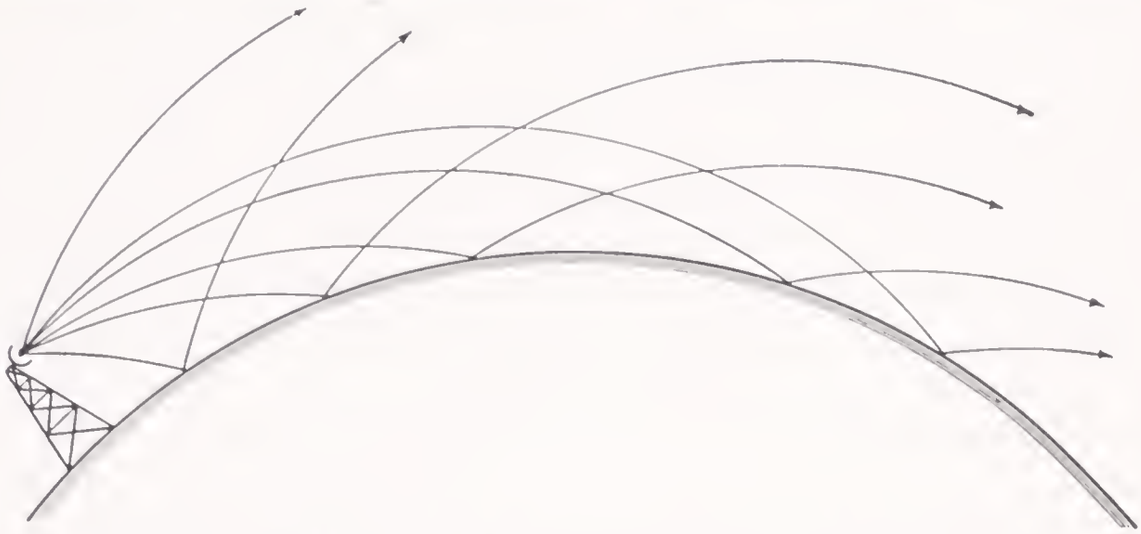


FIG. 14. Surface trapping resulting from nonstandard refraction.

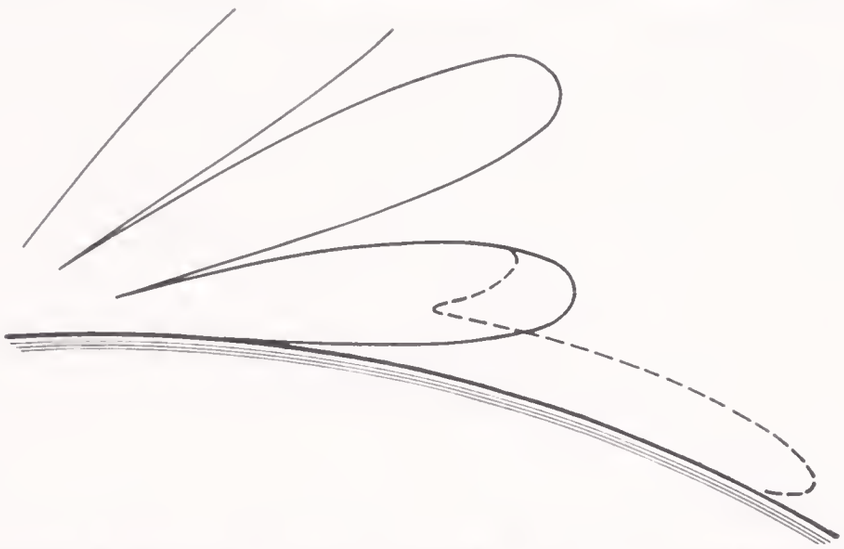


FIG. 15. Increased low-angle and surface coverage caused by surface trapping of energy.

the curvature of the earth; hence some of the energy is trapped and tends to follow the surface of the earth, as shown in Fig. 14. Only the rays which leave the antenna at angles within about a degree of the horizontal direction are affected. High-angle radiation is not bent sufficiently to be trapped.

The effect of this trapping on the radar-coverage pattern is to extend the lobe or lobes which lie within the region in which the energy is trapped, as shown in Fig. 15.

There are, of course, many variations in the effects on propagation produced by nonstandard refraction. Extensive studies have been made in an attempt to relate weather conditions to radar performance, but at

present no reliable method is available for accurately predicting radar performance under widely varying conditions. Only rough estimates of performance are possible under a few rather idealized nonstandard conditions.[1]

9. Atmospheric Reflections. When a radar antenna is in the normal search position, atmospheric disturbances such as rain, storms, moisture-laden clouds, or cold fronts may be detected to ranges of 150 miles or more. These echoes tend to have a more fuzzy appearance than do normal radar echoes. A weather disturbance that gives a radar echo can be tracked to determine its direction of motion and speed, so that the nature of the disturbances can be fairly definitely established. Thus radar is of assistance in weather forecasting. In connection with meteorological observations, radar is also used to track balloons and so to determine local wind velocity at various altitudes. In this application, as in other tracking problems, radar excels optical methods in that greater ranges are obtainable and tracking can be accomplished at any time of day or night and under any visibility conditions, provided the echoes from an atmospheric disturbance do not mask the echo from the object being tracked.

Some evidence also exists of echoes being received from the atmosphere in absence of precipitation. The origin of these echoes, often referred to as "angels," is not completely understood. Conceivably such signals might arise from the presence of birds, from swarms of insects, or perhaps from irregularities in the refractive index.

10. Character of Radar Echoes. In the preceding article, it was mentioned that echoes from atmospheric disturbances are not so sharply defined as echoes from other objects. This is to be expected, since the outline of a cloud or storm is itself not so sharply defined as that of an aircraft, a ship, or a strip of land. However, even the latter objects do not reflect a steady signal.

One method used in studying the character of radar echoes employs high-speed photography. Photographs of an indicator screen are made with a high-speed motion camera, the camera and the radar transmitter being synchronized so that each range trace appears on a separate frame. The pictures may be viewed individually or projected at a slow speed, and thus rapid variations of intensity may be followed.

A summary of the results of such studies of radar echoes is as follows:

1. The most rapid random variations of all occur in storms. The appearance of storm echoes on the indicator is similar to the appearance of receiver noise.

[1] For a summary of the problems involved in forecasting radar performance, see "Variations in Radar Coverage," JANP 101, (Joint Communications Board, Washington, D.C., 1945); and "Weather Forecasting for Radar Operations," Weather Division Report No. 614 (Headquarters Army Air Forces, March, 1944).

2. Random variations also occur from sea return, which is the name given to the echoes obtained from the surface of the sea. Since this return consists of the echoes from countless individual sources, the positions of which are continually changing, the random nature of this type of echo is understandable.

3. The amplitude of the echo signal from an aircraft fluctuates violently and rapidly because the aspect of the aircraft relative to the radar set is continually changing and because much of the return comes from the rotating propellers. Superimposed on the rapid variations are somewhat slower variations due to the lobe structure of the vertical-coverage pattern of the radar set.

4. The echo from a ship fluctuates much less rapidly than that from an airplane, since both the aspect of the ship and the range change slowly.

5. The echo from a rocky land mass is fairly steady, but the echo from a wooded terrain fluctuates considerably. The amount and rate of fluctuation of the echo from wooded terrain are related to wind velocity and radar wavelength.

6. A single isolated object, such as a tower, produces a fairly steady signal, but a slight motion of the target causes some variation in signal strength, particularly with microwave radar. If propagation takes place over water, the tides may cause a slow regular variation in echo signals. At long ranges, variations in atmospheric refraction also produce changes in echo-signal strength.

INDEX

A

Altimeter, radio, 44
Ampere's law, 536
Amplidyne, 316–319
Amplifiers, 56, 186–200, 307, 339, 364, 401–432, 873
  i-f, 412–432
    double-tuned, 419
    input stages for, 427
    single-tuned, 412
    stagger-tuned, 424
  linear, 364
  linear-logarithmic, 465
  overdriven, 56–61
  power, 186–200
  r-f, 339, 872
  servo, 307–309
  video, 401–412
    blocking in, 403
    compensated, 409–412
    gain-bandwidth product, 408
    half-power frequencies, 407
Angels, 976
Antenna arrays, beam angle of, 933
  broadside, 926
  elimination of back radiation from, 929–933
Antenna beam angle, 11, 912, 933
Antenna feeds, 8, 950–957
  coaxial-line, 952
  Cutler, 956
  front waveguide, 954
  rear waveguide, 956
  sectoral horn, 955
Antenna lobes, 11, 947
Antennas, 8, 9, 33, 34, 895–958
  cosecant-squared, 957
  directional, 7
  half-wave, 895
    patterns of, 910–913
    radiation fields, 903–905
      analysis of, 910
  input-impedance of, 899–903
Antennas, parabolic, 942–950
    feeds for, 950
      (*See also* Antenna feeds)
  parasitic, 937–942
  radiation patterns of, 913–924
  Yagi, 941
Anti-interference circuits, 461–467
Anti-t-r devices, 46, 854, 868, 870
Aperture coupling, 669–689
Applegate diagram (*see* Klystrons)
Aquadag, 218
Astigmatism, cathode-ray-tube, 218
Atmospheric reflections, 976
Attenuation constant, 517, 541
Attenuators, 840–849
  coaxial-line, 841
  cut-off, 847
  waveguide, 843
Automatic back bias, 463
Automatic frequency control, 49
Automatic frequency-control circuits, 452–461
Automatic gain control, instantaneous, 463
Autotransformer, electrical-zero, 298

B

Beacons, radar, 44
Beam angle, 11, 912, 933
Bends, waveguide, 809, 816
Bias, signal, 82–85
Bias circuits, 73–75
Blocking oscillator, 134–138
  driven, 134
  free-running, 138
  line-controlled, 204
Bootstrap principle, 201
  (*See also* Modulator driver circuits; Sweep circuits)
Bunching, 747, 751, 779–781

C

Cathode-coupled multivibrator, 99
Cathode followers, 61–66

Cathode-ray tubes, 215–227
designations and ratings, 225
electrostatic, 216
astigmatism, 218
deflection factor, 217
defocusing effects, 218
electron gun, 216
fluorescent screens for, 220, 224
magnetic, 221
deflection yoke, 223
electron gun, 222
Cathode-return circuit, 716
Cavity resonators (*see* Resonators)
Characteristic resistance, 142, 471
Choke, r-f, 809, 851
Choke joints, 809–812
Clampers, 82–85, 302
Clipping circuits, 54–66
with amplifiers, 56
with cathode followers, 61
with diodes, 54
Coaxial cables, properties of, 544
Complex numbers, 349
Conical scan, 30, 952
Coupling circuits, 79–82, 131
Coupling loops, 669, 674
Coupling probes, 668, 682
Coupling reactance, 646
Coverage diagrams, 971–974
Crystal mixers (*see* Mixers)
Crystal shutters, 858
Crystals, silicon, 37, 371–376
cartridges, 372
spreading resistance of, 373
type numbers, 376
Current density, linear, 548
Cutler feed, 956

D

Decibel, 396
Decoupling filters, 73–75
Delay line, 150
Detectors, 432–441
diode, 432
triode and pentode, 439–441
infinite-impedance, 440
Dielectric constant, 534
Differentiating circuit (*see* Peakers)
Diode, gas, 167
pick-off, 86
D-c restorers (clampers), 82–85, 302
D-c servo motor, 309
Directional couplers, 810, 834–839
Discriminator, 453
Disk-seal triodes, 35, 699
Divergence, 967
Dots, significance of, on diagrams, 56, 124
Duplexer, 10
Duty ratio, definition of, 4

E

Earth reflections, effect on range, 965–967
Echo box, 26, 805, 890
ringing time, 892
Echoes, radar, character of, 976
Electric-field intensity, 532
Electric flux, 532
Electric-flux density, 532
Electrical-zero autotransformer, 298
Electromagnetic waves, basic concepts of, 547
plane transverse, reflection of, normal, 556
oblique, 560
polarization of, 555
types of, 554–556
Electron gun, 216, 222
Exponential function, 68

F

F factors, 918, 924–929, 962
Faraday's law, 535
Feedback, negative, in servomechanism, 329
Feeds, antenna (*see* Antenna feeds)
Filters, decoupling, 73–75
Fluorescent screens, 220, 224
Fourier analysis, 341
Frequency bands, radar, 33

G

G factors, 922, 929–933, 962
Grid loading, 367–371
Grid-return circuits, analysis of, 716–720
Ground wave, 960
Group velocity, 577

H

Horizon, radar, 968
Hunting in servomechanisms, 325
Hybrid junctions, 825–834, 883

I

Image-frequency response of receivers, 398
Impedance-changing devices, 523–530, 601–608, 831, 855
  bandwidth of, 530
  (*See also* Irises; Transformers, impedance)
Indicators, 46, 215–276
  (*See also* Cathode-ray tubes; Sweep-current circuits; Sweep-voltage circuits)
Induced grid noise, 388
Induction motor, two-phase, 319, 322
Inductors, saturable-core, 168
Instantaneous automatic gain control, 463
Inverse-fourth-power law, 21
Irises, 601, 604–608, 810, 856

J

Junctions, hybrid, 825–834, 883
  magic-T, 531–534
  waveguide-to-coaxial line, 684

K

K band, 33
Keep-alive electrode in t-r tubes, 858
Keeper, 805
Klystrons, 37, 745–783
  reflex, 750
    Applegate diagram for, 751
    bunching in, 751
    characteristics, 753, 755–760
    hysteresis and load effects, 761
    tuning, 760
    types, 753
    voltage modes, 753
  two-resonator, 745–750
    Applegate diagram for, 747
    bunching in, 747
    drift space, 746

L

L band, 33
Lead networks, 331–336
Lighthouse-tube oscillator, 722–874
Lighthouse tubes, 699, 872
  (*See also* Disk-seal triodes)
Line stretcher, 853
Linear current density, 548
Lobe switching, 30
Lobes, antenna, 11, 947
Local oscillators, 36, 48, 340, 699, 745, 874, 878
  (*See also* Klystrons; Tubes, ultrahigh-frequency)
Loop coupling, 669, 674
Loran, 46

M

Magic T, 531–534
Magnetic-field intensity, 535
Magnetic-flux density, 535
Magnetron characteristics, 781–792
  mode-line diagram, 782
  mode shift, 791
  mode skip, 791
Magnetron resonators, 769–776
  methods of strapping, 774
  mode number, 770
  modes, 770–773
    (*See also* Magnetron characteristics)
Magnetron spectra, examples of, 805
Magnetrons, 35, 763–805
  construction, 763–765
  electron motion in, 776–781
    bunching, 779–781
    phase-focusing effect, 781
  frequency pulling, 802
  frequency splitting, 802, 804
  operation and care, 803–805
  performance charts, 792–795
  pulling figure, 799
  Rieke diagrams, 795–799
  rising-sun, 768
  straps, 764
    (*See also* Magnetron resonators)
  type numbers, 765–769
Master-oscillator radar system, 50
Matching devices (*see* Impedance-changing devices; Transmission lines)

Matching posts, 601, 602
Miller effect, 230
Mixers, balanced, 871–886
  conversion loss in, 445
  crystal, 441–449
  single-ended, 876–881
  vacuum-tube, 449–452
    conversion transconductance of, 451
Modulator driver circuits, 200–214
  bootstrap, 200
  delay-line controlled, 209
  line-controlled blocking-oscillator, 204
  shorted-end line, 210
Modulators, 153–214
  blocking-oscillator, 214
  driver-power-amplifier, 186–214
    drivers (*see* Modulator driver circuits)
  line-pulsing, 169–186
    charging methods for, 170–176
    pulse-forming networks for, 160–166
  rotary-spark-gap, 176
    with saturable-core reactor, 183
  switching devices, 166–169
  types of, 153
Motors, 309–312, 319
  d-c servo, 309
  two-phase induction, 319, 322
Multivibrators, classification of, 92
  free-running, 92–97
  one-cycle, 97–104
    cathode-coupled, 99
    plate-coupled, 97

N

Neper, definition of, 514
Networks, pulse-forming (*see* Pulse-forming networks)
  resonant modes of, 613
Noise, 16, 376–396
  antenna, 383
  characteristics of, 376–379
  crystal, 384–386
  interception, 391
  in linear passive circuits, 379
  in receivers, 376
  shot-effect, 385
  thermal agitation, 384–386
  vacuum-tube, 384–386
  white, 381
Noise figure, 392
Noise resistance, equivalent, 389
Noise-temperature ratio, 385

O

Open-end line, charging of, 154–158
  pulse formation by discharge of, 158
Oscillations, parasitic, 709
  in servomechanisms, 324–328
Oscillators, 134, 204, 248, 697–744
  blocking, 134–138, 204
  Colpitts, 707
  crystal-controlled, 248–252
  high-frequency triode, 697–744
    grid-return, 720–724
    loading, 724–727, 734–737
    multitube, 712
    output circuits for, 727–734
    plate-return, 709
    pulse, 740–744
    ring, 713
    self-pulsed, 739–741
    tubes for, 698
    ultra-audion, 706–709

P

Parabolic reflectors, 943
Parasitic oscillations, 709
Parasitic suppressors, 709
Peakers, *R-C*, 75–79
  R-L-C, 122
Permittivity, 533
Phantastron delay circuit, 104–110
Phase constant, 474
Phase-inverter circuits, 234–239
Phase-sensitive rectifiers, 312–316
  keyed, 315
  with load, in cathode circuit, 314
    in plate circuit, 312
  using diodes, 315
Phase velocity, 577
Phosphors, 220, 224, 225
Plan-position indication, 12
Plan-position indicator, fixed-coil, 271, 302–305
  rotary-coil, 267
Plate resistance, diode, 55
  triode, 60, 61
Plate-return circuit, 716

Polarization of electromagnetic waves, 555
Posts, matching, 601, 602
Potentiometer, tapered, 267
Poynting vector, 553
Pre-t-r tube, 868
Presentation types, 226
Probe coupling, 668, 682
Propagation, 959–977
  over earth, 965
  factors influencing, 959
  line-of-sight, 959
  over sea water, 964
Propagation constant, 518
Pulling figure, magnetron, 799
Pulse formation, by discharge of open-end line, 158
  by shorted-end line, 159
Pulse-forming networks, 160–169
  characteristics of loads for, 164
  switching devices for discharging, 166
Pulse repetition frequency, 3
Pulse transformers (*see* Transformers)
Pulses, carrier-frequency, reproduction of, 359
    sine-wave components of, 346
  video, frequency spectrum of, 344
    reproduction of, 356
    sine-wave components of, 343

Q

Q (quality factor), 116, 617, 624, 661
  loaded, 661
  of parallel *R-L-C* circuit, 116
  unloaded, 661

R

R-C circuits, 66–114
  effect of tube and wiring capacitances in, 110
  sine-wave response of, 349–352
  transients in, 67
R-C peakers, 75–79
R-L circuits, 114
R-L-C circuits, 114–152, 352–356
  parallel, 115, 352
    damping in, 119
    resonant frequency of, 116
    sine-wave response of, 352
R-L-C circuits, parallel, transients in, 115
  series, 117
R-L-C peakers, 122
Radar, advantages of, 1
  definition of, 1
  limitations, 23
  pulsed, basic principles of, 3
Radar bands, designations of, 33
Radar echoes, 976
Radar equipment, types of, 39
Radar frequencies, 32
Radar horizon, 968
Radar range, effect of earth reflections on, 965
Radar-range equation, 18
Radar set, block diagram of, 46
Radiation, components of, 960
R-f choke, 809, 851
  (*See also* Choke joints)
R-f envelope indicator, 804
R-f loads, 845–847
R-f system components, 806–810
R-f transmission lines (*see* Transmission lines)
R-f transmitting and receiving systems, 806–894
Range, determination of, 2
Range accuracy, 28
Range-height indication, 227
Range-marker circuit, 123
Range-step generator, 86–92
Range-time conversion factors, 6
Receiver frequencies and bandwidths, 396
Receiver protective device, 10
Receivers, 48, 338–467, 871
  block diagrams, 339
  component circuits of, 401–467
    anti-interference, 461–467
    automatic frequency control, 49
    (*See also* Amplifiers; Detectors; Mixers)
  general principles of, 338
  image-frequency response of, 398
  ultrahigh-frequency, input circuits of, 871–876
Reference directions, explanation of, 56
Reflection, law of, 560
  from a perfectly conducting plane, 961
Reflection coefficient, of sea water, 965
  transmission-line, 486, 495
    graphical determination of, 493

Reflections, atmospheric, 976
  earth, effect of, on radar range, 965–967
Refraction, nonstandard atmospheric, 974
  standard, 968
Resolution, angular, 10, 15
  range, 26
Resonant lines and cavities, 613–696
Resonators, 35
  cavity, 35, 625
    circular-cylinder, 637
    re-entrant, 636
  coaxial-line, 633
  coupling devices, 668–670
  one-entry, 639
    coupling to, 645
  rectangular-waveguide, 628
  transmission-line, 619–625, 649
    coupling to, 654
  two-entry, 665
Right-hand-screw rule, 536
Ringing circuit, 120
  for Type J presentation, 244–248
Rotary joints, 809, 812–816

S

S band, 33
Saturable-core inductors, 168
Saw-tooth generator, 69
Screws, matching, 602
Self-synchronous radar system, 50
Sensitivity-time control, 465
Servo motor, d-c, 309
Servo systems, components of, 306
Servo using Ward Leonard control system, 308
Servomechanisms, 305–337
  antihunt devices for, 328
  applications of, 336
  hunting in, 325
  lead networks for, 331
  motor, 322
  negative feedback in, 329
  oscillations in, sustained, 325
    transient, 324
  using two-phase induction motor, 322
Signal bias, 82–85
Signal integration, 467
Signal-to-noise ratio, 18
Sine-wave analysis, 341
Skin depth, 540
Skin effect, 539
Sky wave, 960
Smith chart, 500
Spectrum analyzer, 681, 805, 893
Standing-wave indicators, 886–889
Standing-wave ratio, 489
Stub supports, coaxial line, 820–823
Sweep circuits, 227–276
  bootstrap, 231
  delayed and expanded, 239
  feedback-amplifier, 262
Sweep-current circuits, 252–276
Sweep-voltage circuits, 227–252
  methods of improving linearity of, 229
Synchro, 277–305
  applications of, 300
  designations of, 295
  differential, 292
  oscillation dampers for, 288
Synchro capacitors, 296
Synchro control transformer, 289
Synchro generator, 277, 280
  electrical zero of, 282
Synchro motor, operation of, 277, 284
Synchro systems, geared, 298
Synchronizing voltage, 6

T

T junctions, 818–825
  coaxial-line, 672, 820
  $\mathscr{E}$-plane, 823–825
  $\mathscr{H}$-plane, 823–825
$TE_{1,0}$ wave, 570
  characteristic wave impedance of, 598
  guide wavelength, 574
  pictorial representation of, 573
Temperature inversion, 974
Test equipment, r-f, 886–894
Thyratrons, argon, 168
  hydrogen, 168
Time constant, 68
Timing circuits, 54–152
Tracking, types of, 43
Transformers, balanced-to-unbalanced, 849–851
  ideal, 125
  impedance, 523–530, 855, 856
    double-stub, 525–528, 855
    quarter-wave, 528–530, 856

Transformers, impedance, single-stub, 523–525, 855
pulse, 123
with bifilar winding, 179
as coupling devices, 131
equivalent circuits of, 124
resonant, 666
Transit-time effects, 369
Transmission-line charts, 499–505
Transmission-line parameters, 531–544
Transmission-line resonators (*see* Resonators, transmission-line)
Transmission lines, 46
artificial, 139, 140
applications of, 150
cutoff frequency, 140
traveling waves on, 139
coaxial, 34, 46, 533, 538, 541, 542
parallel-strip, 547, 550–552
parallel-wire, 34, 534, 539, 541, 542
r-f, 468–544
characteristic resistance, 471, 541
efficiency, 520
impedance-changing devices, 523–528
input admittance, 505
input impedance, 482, 495, 521
losses and attenuation, 514
operation of, 506
power and power capacity, 510–514
standing waves on, 479
switching operations, 143
transient and steady-state operation, 506
types, 495, 521
terminated by resonators, 658
traveling waves on, 471
T-r (transmit-receive) device, 10
T-r tubes, 857–860, 871
T-r systems, ultrahigh-frequency, 860–863
coaxial line microwave, 863–866
waveguide, 866–871
Transmitters (*see* Magnetrons; Oscillators)
Transponders, 45
Trapezoidal generator, 258
Tubes, linear equivalent circuits for, 364
microwave local-oscillator, 753
power-amplifier, for modulators, 196
ultrahigh-frequency, 697–703
Type A indicator, basic circuit of, 227
Type A presentation, 12
(*See also* Presentation, types of)
Type B presentation, 12
azimuth-sweep circuit for, 265
Type J presentation, 242
crystal-controlled oscillator for, 248
ringing circuit for, 244
Type P presentation, 12

V

Video amplifiers (*see* Amplifiers)
Video pulses, amplification of, 404

W

Waveguides, 35, 545–612
applications of, 545
circular, 608–612
corners, 809, 817
fields for frequencies below cutoff, 591
impedances, 596, 599
rectangular, 570
attenuation constant, 585
dimensions of, 590, 591
dominant wave, 545, 590
(*See also* $TE_{1,0}$ wave)
group velocity, 577
higher order wave types, 587
impedance-changing devices, 601–608
phase velocity, 577
power and attenuation in, 580
twists, 809, 818
Wavelength, 474
Wavemeters, 677, 889
one-entry, 678
two-entry, 680

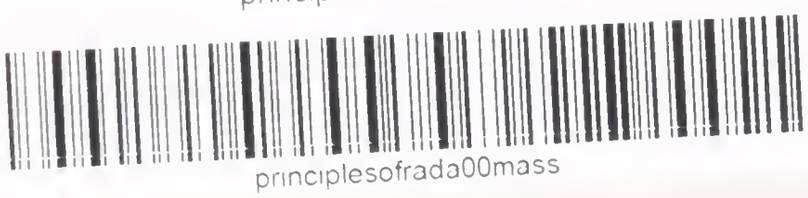
X

X band, 33

Y

Yagi antenna, 941

Z

z-plane circle diagram, 503

principlesofrada00mass